Applying clique-decomposition for computing Gromov hyperbolicity

Nathann Cohen, David Coudert, Guillaume Ducoffe, Aurélien Lancin

Applying clique-decomposition for computing Gromov hyperbolicity

Nathann Cohen, David Coudert, Guillaume Ducoffe, Aurélien Lancin

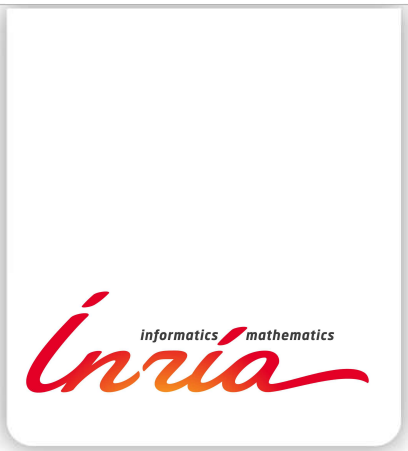
RESEARCH REPORT

N° 8535

May 2014

Project-Team COATI

ISSN 0249-6399 ISRN INRIA/RR--8535--FR+ENG



Applying clique-decomposition for computing Gromov hyperbolicity*

Nathann Cohen†, David Coudert‡§, Guillaume Ducoffe¶, Aurélien Lancin‡§

Project-Team COATI

Research Report n° 8535 — version 2 — initial version May 2014 — revised version July 2014 — 33 pages

Abstract: The shortest-path metric d of a connected graph G is *δ-hyperbolic* if, and only if, it satisfies $\mathrm{d}(u,v)+\mathrm{d}(x,y) \leq \max\{\mathrm{d}(u,x)+\mathrm{d}(v,y), \mathrm{d}(u,y)+\mathrm{d}(v,x)\}+2\delta$, for every 4-tuple u,x,v,y of G. We investigate some relations between the hyperbolicity of a graph and the hyperbolicity of its *atoms*, that are the subgraphs resulting from the clique-decomposition invented by Tarjan [34, 45]. More precisely, we prove that the maximum hyperbolicity taken over all the atoms is at least the hyperbolicity of G *minus one*. We also give an algorithm to slightly modify the atoms, which is at no extra cost than computing the atoms themselves, and so that the maximum hyperbolicity taken over all the resulting graphs is *exactly* the hyperbolicity of G. An experimental evaluation of our methodology is provided for large collaboration networks. Finally, we deduce from our theoretical results the first *linear-time* algorithm to compute the hyperbolicity of an outerplanar graph.

Key-words: Hyperbolicity, algorithm, graph, decomposition.

* This work has been partially supported by ANR project Stint under reference ANR-13-BS02-0007, ANR program "Investments for the Future" under reference ANR-11-LABX-0031-01, and by European project FP7 EULER (Grant No.258307).

† LRI, Laboratoire de Recherche en Informatique, Université Paris-Sud 11, France.
‡ Inria, France
§ Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
¶ ENS Cachan, 61 Avenue du Président Wilson, 94230 Cachan, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Décomposition par des cliques-séparatrices pour le calcul de l'hyperbolicité de Gromov

Résumé : La métrique d des plus courts chemins d'un graphe connexe G est *δ-hyperbolique* si, et seulement si, elle satisfait $\mathrm{d}(u,v)+\mathrm{d}(x,y) \leq \max\{\mathrm{d}(u,x)+\mathrm{d}(v,y), \mathrm{d}(u,y)+\mathrm{d}(v,x)\}+2\delta$, pour tout quadruplet u,x,v,y de G. Nous étudions la relation entre l'hyperbolicité d'un graphe et celle de chacun de ses *atomes*. Ces derniers sont les sous-graphes résultant de la décomposition d'un graphe par des cliques-séparatrices [34, 45]. Plus précisemment, nous montrons que l'hyperbolicité d'un atome est au plus l'hyperbolicité de G *moins un*. Nous proposons un algorithme pour modifier les atomes de sorte que la valeur maximale de l'hyperbolicité de ces atomes modifiés soit exactement l'hyperbolicité de G. La complexité de cet algorithme est la même que celle de la décomposition du graphe par des cliques-séparatrices. Nous évaluons expériementalement cette méthode sur des graphes de collaborations (co-auteurs). Enfin, nous proposons un algorithme pour calculer en temps linéaire l'hyperbolicité des graphes planaires extérieurs.

Mots-clés : Hyperbolicité, algorithme, graphe, décomposition.

Contents

1 Introduction

In this paper, we primarily aim to study graph hyperbolicity, from an algorithmic point of view. This parameter was first introduced by Gromov in the context of automatic groups [26], then extended to more general metric spaces, including the shortest-path metrics of simple graphs. Hyperbolicity of unweighted graphs, and its tight relations to other metric parameters, has received growing attention over the last decades, especially due to its practical applications in approximation algorithms [13], routing [9], network security [29] and bioinformatics [12, 22], to name a few. The reader may refer to [1, 21] for a recent survey.

However, the computational cost of hyperbolicity has received less attention. So far, the best-known algorithm for computing the hyperbolicity of a graph [23] is impractical for large-scale graphs such as the graph of the Autonomous Systems of the Internet, road maps, etc. This both comes from its challenging implementation, relying on fast square matrix multiplications, and its time complexity which is strictly more than cubic. While practical advances have been made, improving the computational cost on certain graph classes (see [14, 31]), a recent theoretical work [15] suggests that an algorithm for the problem with a significant speed-up for *all graphs* is unlikely to exist. This motivates the study of structural properties that may help to decrease the running time of the computation of hyperbolicity.

Our approach relies on divide-and-conquer techniques, especially *graph decompositions*. Roughly, we aim to reduce the computation of the hyperbolicity of a graph to the computation of the hyperbolicity of (some of) its subgraphs, so that we can decrease the size of the input graphs to deal with. A first step toward this direction was the result in [41], where the author proved that the hyperbolicity of a connected graph is the maximum hyperbolicity taken over all the subgraphs obtained either via the modular decomposition [24], or with the split-decomposition [16].

We here address a similar question for the subgraphs obtained via the *clique-decomposition*, also known as *atoms*. This decomposition was first introduced by Tarjan in [45], then made unique by Leimer in [34]. On the theoretical side, it is already known that clique-decomposition can be applied to speed-up the computation of many graph parameters, including metric parameters that are related to hyperbolicity, such as tree-length [20]. Also, complex networks are expected to be decomposable w.r.t. clique-decomposition, especially when they are related to phylogenetical data, text-data mining, or distance data [6, 18, 30]. This makes our approach practical for large-scale graphs.

We will prove that while the hyperbolicity of a graph cannot be deduced from the hyperbolicity of its atoms (Section 3), it yields an approximation with additive constant 1 of this parameter (Section 4). Additionally, we characterize the cases for which this small additive distortion might happen. In Section 5, we will show how each atom can be modified in order to compute exactly the hyperbolicity, and provide a complexity analysis of the procedure. Experiments in Section 7 show the benefit of our method in terms of size of the graph, when applied to some real collaboration networks.

Finally, we will see in Section 6 that our method is beneficial for the class of outerplanar graphs, as it gives a linear-time algorithm to compute the hyperbolicity of these graphs.

Definitions and notations used in this paper will be introduced in Section 2.

2 Definitions and notations

We essentially rely on the graph terminology of [10, 19]. All graphs considered in this paper are finite, unweighted and simple. We here only emphasize some notions related to metric graph

theory. The reader may refer to [3] for a survey about this domain. Given two vertices u, v, a *uv-path* of length $l \geq 0$ is a sequence of vertices $(u = v_0 v_1 \dots v_l = v)$, such that $\{v_i, v_{i+1}\}$ is an edge for every i. In particular, G is *connected* if there exists a uv-path for all pair $u, v \in V$, and in such a case the *distance* $\mathrm{d}_G(u, v)$ is defined as the minimum length of a uv-path in G. Note that it yields a discrete metric space (V, d_G), also known as the shortest-path metric space of G. We will write d instead of d_G whenever G is clear from the context, and we denote by $\mathrm{d}(u, X) = \min_{x \in X} \mathrm{d}(u, x)$ the distance between a vertex u and a set X of vertices.

Our proofs use the notions of subgraphs, *induced* subgraphs, as well as *isometric subgraphs*, the latter denoting a subgraph H of a graph G such that $\mathrm{d}_H(u, v) = \mathrm{d}_G(u, v)$ for any two vertices $u, v \in H$.

δ-hyperbolic graphs. Graph hyperbolicity provides tight bounds on the worst additive distortion of the distances in a (connected) graph when its vertices are embedded into a weighted tree. Several definitions exist, some of them considering graph metrics that are slightly different from the usual shortest-path metric, but they are equivalent to it up to a linear-function [5, 17, 26]. Moreover, 0-hyperbolic graphs are exactly the connected graphs whose given metric is a *tree metric*, which makes hyperbolicity a tree-likeness parameter. Especially, the shortest-path metric of a graph is 0-hyperbolic if, and only if, it is a *block-graph*, that is a graph whose biconnected components are complete subgraphs [4, 28]. This class of graphs includes trees and cliques, and a block-graph can be recognized in linear $O(n + m)$-time.

Definition 1 (**4-points Condition, [26]**). *Let G be a connected graph. For every 4-tuple u, x, v, y of vertices of G, we define $\delta(u, v, x, y)$ as half of the difference between the two largest sums amongst*

$$S_1 = \mathrm{d}(u, v) + \mathrm{d}(x, y),\ S_2 = \mathrm{d}(u, x) + \mathrm{d}(v, y) \text{ and } S_3 = \mathrm{d}(u, y) + \mathrm{d}(v, x)$$

The hyperbolicity of G, denoted by $\delta(G)$, is equal to $\max_{u,x,v,y \in V(G)} \delta(u, v, x, y)$. Moreover, we say that G is δ-hyperbolic*, for every $\delta \geq \delta(G)$.*

It is straightforward, by the above definition, to compute graph hyperbolicity in $\theta(n^4)$-time. Currently, the best-known *theoretical* algorithm for the problem runs in $O(n^{3.69})$-time [23], and the best-known *practical* algorithm has $O(n^4)$-time complexity [14]. Recognizing graphs with small hyperbolicity upper-bounded by $\frac{1}{2}$ is computationally equivalent to decide whether there is a chordless cycle of length 4 in a graph, and it can be done in $O(n^{3.26})$-time by using fast rectangular matrix multiplication [15, 33]. Note also that the hyperbolicity of a connected graph is the maximum hyperbolicity taken over all its biconnected components.

Atoms and clique-separators. Given a connected graph $G = (V, E)$, we name *separator* a subset of vertices $X \subset V$ such that the removal of X disconnects the graph. When the induced subgraph $G[X]$ is a complete graph, then we say that X is a *clique-separator*. More generally, we say that X is a clique-separator of $S \subseteq V$ if S intersects (at least) two distinct connected components of $G \backslash X$, and S is then said to be *separable*. The clique-decomposition of a graph (see Figure 1) is the collection of all its maximal non-separable sets of vertices (we will call them *atoms*) [45]. The decomposition is unique [34], and it can be computed in $O(nm)$-time [34, 45].

Notations. Let us fix some notations for the proofs. Given two subsets A, B of vertices, we say that a separator X is an $(A|B)$-separator if it disconnects any $a \in A \backslash X$ from any $b \in B \backslash X$. In particular, any separator X containing A or B is an $(A|B)$-separator, and in such a case we call X a *trivial* $(A|B)$-separator.

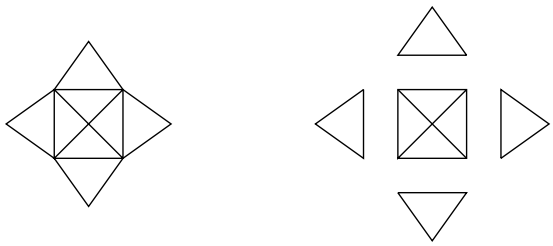


Figure 1: Clique-decomposition of a graph into five atoms.

We also denote by $(a|b_1, b_2, b_3)$ a 4-tuple such that $a \in A$ and $b_1, b_2, b_3 \in B$. In the same way, we denote by $(a_1, a_2|b_1, b_2)$ a 4-tuple such that $a_1, a_2 \in A$ and $b_1, b_2 \in B$.

Finally, a clique-separator is called a *clique-minimal separator* if there exists some pair u, v such that it is an inclusionwise minimal, non-trivial $(u|v)$-clique-separator. It has to be noted, perhaps counter-intuitively, that a clique-minimal separator may be contained into another one.

3 Hyperbolicity and clique-separators

We analyze in this section the relationship between the hyperbolicity of a graph and a given clique-separator, leading to the approximation with additive constant of Theorem 11. It begins with an observation about $(a_1, a_2|b_1, b_2)$ 4-tuples and the diameter $\mathrm{diam}(X) = \max_{u,v \in X} \mathrm{d}_G(u, v)$ of an $(A|B)$-separator X.

3.1 Hyperbolicity of $(a_1, a_2|b_1, b_2)$ 4-tuples

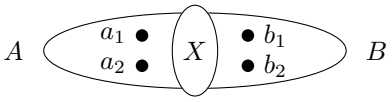


Figure 2: Illustration of an $(A|B)$-separator.

Lemma 2. *Let X be a an $(A|B)$-separator of a connected graph G. For every $(a_1, a_2|b_1, b_2)$ 4-tuple, we have $\delta(a_1, a_2, b_1, b_2) \leq \mathrm{diam}(X)$.*

Proof. Recall that we have:

$$\begin{aligned} S_1 &= \mathrm{d}(a_1, a_2) + \mathrm{d}(b_1, b_2), \\ S_2 &= \mathrm{d}(a_1, b_1) + \mathrm{d}(a_2, b_2), \\ S_3 &= \mathrm{d}(a_1, b_2) + \mathrm{d}(a_2, b_1). \end{aligned}$$

We can assume without loss of generality (w.l.o.g.) that $S_2 \geq S_3$. Writing $\mathrm{d}(v, X) = \min_{x \in X} \mathrm{d}(v, x)$, we also know that for all $i, j \in \{1, 2\}$:

$$\begin{aligned} \mathrm{d}(a_i, b_i) &\geq \mathrm{d}(a_i, X) + \mathrm{d}(b_i, X) \\ \mathrm{d}(a_i, a_j) &\leq \mathrm{d}(a_i, X) + \mathrm{diam}(X) + \mathrm{d}(a_j, X) \end{aligned}$$

If $S_1 \geq S_2$, we get:

$$\begin{aligned} S_2 =& \mathrm{d}(a_1, b_1) + \mathrm{d}(a_2, b_2) \\ \geq& \mathrm{d}(a_1, X) + \mathrm{d}(b_1, X) + \mathrm{d}(a_2, X) + \mathrm{d}(b_2, X) \\ \geq& \big[\mathrm{d}(a_1, X) + \mathrm{d}(a_2, X) + \mathrm{diam}(X)\big] + \\ & \big[\mathrm{d}(b_1, X) + \mathrm{d}(b_2, X) + \mathrm{diam}(X)\big] - 2\,\mathrm{diam}(X) \\ \geq& S_1 - 2\,\mathrm{diam}(X) \end{aligned}$$

Hence we have that $\delta(a_1, a_2, b_1, b_2) \leq (S_1 - S_2)/2 \leq \mathrm{diam}(X)$. It can be shown similarly that $S_3 \geq S_2 - 2\,\mathrm{diam}(X)$, and when $S_2 \geq S_1$ that $S_1 \geq S_2 - 2\,\mathrm{diam}(X)$. Consequently, $\delta(a_1, a_2, b_1, b_2) \leq \mathrm{diam}(X)$ in all cases. □

Corollary 3. *Let X be a an $(A|B)$-clique-separator of a connected graph G. For every $(a_1, a_2|b_1, b_2)$ 4-tuple, we have $\delta(a_1, a_2, b_1, b_2) \leq 1$.*

3.2 Hyperbolicity of $(a|b_1, b_2, b_3)$ 4-tuples

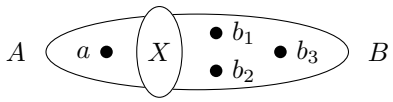


Figure 3: Illustration of an $(a|b_1, b_2, b_3)$-separator.

Note that X being a clique, each vertex $a \in A$ is at distance at least $\mathrm{d}(u, X)$ and at most $\mathrm{d}(u, X) + 1$ from any vertex of X. We now show how this can be used with respect to the hyperbolicity.

Lemma 4. *Let X be an $(A|B)$-clique-separator of a connected graph G, and let $a \in A$. If we add to G a vertex a^* adjacent to $\{x \in X : \mathrm{d}(a, x) = \mathrm{d}(a, X)\}$, then for every $b_1, b_2, b_3 \in B$ we have $\delta(a, b_1, b_2, b_3) = \delta(a^*, b_1, b_2, b_3)$.*

Proof. $\forall x \in X, \mathrm{d}(a, x) \in \{\mathrm{d}(a, X), \mathrm{d}(a, X) + 1\}$ holds as X is a clique. Consequently, for every $b \in B, \mathrm{d}(a, b) = \mathrm{d}(a^*, b) + \mathrm{d}(a, X) - 1$ and replacing a with a^* does not change the hyperbolicity of a, b_1, b_2, b_3. □

Lemma 5. *Let X be an $(A|B)$-clique-separator of a connected graph G. Given an $(a|b_1, b_2, b_3)$ 4-tuple, let $x \in X$ be such that $\mathrm{d}(a, x) = \mathrm{d}(a, X)$. We have $\delta(a, b_1, b_2, b_3) \leq \delta(x, b_1, b_2, b_3) + 1/2$.*

Proof. Let us assume w.l.o.g. that $S_1 \geq S_2 \geq S_3$, where $S_1 = d(a, b_1) + d(b_2, b_3)$, $S_2 = d(a, b_2) + d(b_1, b_3)$ and $S_3 = d(a, b_3) + d(b_1, b_2)$.

We can assume that every vertex $a \in A$ has been replaced by an equivalent vertex a^* as defined in Lemma 4 and so, that all vertices of A have a neighbor in X. In such a situation, any $b \in B$ satisfies $\mathrm{d}(x, b) \leq \mathrm{d}(a, b) \leq \mathrm{d}(x, b) + 1$. Similarly, any sum $S'_i = \mathrm{d}(x, b_i) + \mathrm{d}(b_j, b_k)$, where $\{j, k\} = \{1, 2, 3\} \setminus \{i\}$, satisfies $S'_i \leq S_i \leq S'_i + 1$. Thus for every $i \in \{2, 3\}$:

$$\begin{aligned} \delta(a, b_1, b_2, b_3) \leq& (S_1 - S_i)/2 \\ \leq& (S'_1 + 1 - S'_i)/2 \\ \leq& (S'_1 - S'_i)/2 + 1/2 \end{aligned}$$

In particular, if $S'_1 \neq \max\{S'_1, S'_2, S'_3\}$, then we have $\delta(a, b_1, b_2, b_3) \leq \frac{1}{2}$ by the choice of $S'_i = \max\{S'_1, S'_2, S'_3\}$. Otherwise, we have $\delta(a, b_1, b_2, b_3) \leq \delta(x, b_1, b_2, b_3) + \frac{1}{2}$ by the choice of $S'_i = \max\{S'_2, S'_3\}$. □

3.3 Separable sets

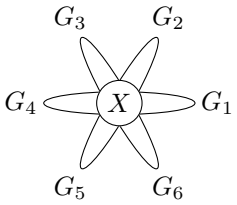


Figure 4: Illustration of separable sets.

Theorem 6. *Let X be a clique-separator of a connected graph G, and let $C_1, \ldots, C_l$ be the connected components of $G \backslash X$. We define $G_i = G[C_i \cup X]$. We have :*

$$\max\{\delta(G_1), \ldots \delta(G_l)\} \leq \delta(G) \leq \max\{\frac{1}{2}, \delta(G_1), \ldots \delta(G_l)\} + 1/2$$

Proof. Note that since a complete subgraph is isometric, then every subgraph G_i is isometric as well. Hence, the lower-bound follows from the four-point definition.

Let us now prove that $\delta(a, b, c, d) \leq \max\{\frac{1}{2}, \delta(G_1), \ldots \delta(G_l)\} + 1/2$ holds for any $a, b, c, d \in V$. We consider a connected component C_i maximizing $|C_i \cap \{a, b, c, d\}|$.

- If $|C_i \cap \{a, b, c, d\}| = 4$ we are done as $\delta(a, b, c, d) \leq \delta(G_i)$.
- If $|C_i \cap \{a, b, c, d\}| = 3$ we can assume that a, b, c, d is an $(a|b_1, b_2, b_3)$ 4-tuple, for the choices of $B = C_i \cup X$ and $A = V \setminus C_i$. By Lemma 5 it follows that $\delta(a, b, c, d) \leq \delta(G_i) + 1/2$.
- Finally, if $|C_i \cap \{a, b, c, d\}| \leq 2$ we can assume that a, b, c, d is an $(a_1, a_2|b_1, b_2)$ 4-tuple, for an appropriate merging of the sets $C_j \cup X$ into two subsets A, B. We know by Corollary 3 that $\delta(a_1, a_2, b_1, b_2) \leq 1$ in this case.

□

The upper-bound of Theorem 6 is actually tight. It can be shown using the graph in Figure 5, constructed from a cycle C_7 of length 7 to which we add a triangle.

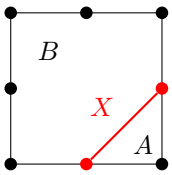


Figure 5: X is an $(A|B)$-clique-separator: we have $\delta(G) = 3/2$, while $\delta(G[B]) = 1$, and $\delta(G[A]) = 0$.

4 Hyperbolicity and clique-decomposition

In Section 3, we explained how a single clique-separator can be used to approximate hyperbolicity. However, this result cannot be used on a whole clique-decomposition as the successive approximations would add up. We thus need to find additional properties to approximate the hyperbolicity of a graph from computations on its atoms in order to prove Theorem 11.

4.1 Relating atoms and 4-tuples with large hyperbolicity

Firstly, we aim to relate to every 4-tuple a, b, c, d with a sufficiently large hyperbolicity, some atom by which all the paths between a, b, c, d go through. Our result involves basic knowledge about *tree-decomposition* (see [8]). We remind the reader that a tree-decomposition of a connected graph G is a tree T whose nodes are labeled by subsets of $V(G)$ (also known as *bags*), and which satisfies the following properties:

- every edge of G is contained into some bag;
- for every vertex $u \in V(G)$, the subgraph induced by the bags containing u is a subtree of T.

Lemma 7. *Let a, b, c, d be a 4-tuple satisfying $\delta(a, b, c, d) \geq \frac{3}{2}$ in a connected graph G. There exists an atom A such that $\forall u \in \{a, b, c, d\} \backslash A$, there is a clique-separator $X_u \subseteq A$ which separates u from $\{a, b, c, d\} \setminus \{u\}$.*

Proof. If the vertices a, b, c, d are contained into a common atom A, then we are done as $\{a, b, c, d\} \setminus A$ is empty. Suppose on the contrary that there is no atom containing the 4-tuple. The authors of [38] proved that there exists a tree-decomposition T_G of G whose bags are exactly the atoms of G. As there is no bag containing a, b, c, d by the hypothesis, there is a unique smallest subtree T of T_G containing a, b, c, d. Note that T is not reduced to a single vertex, and it has at most 4 leaves.

It can be checked that there exists a bag A in T (i.e. an atom of G) such that no connected component of $T \backslash \{A\}$ (and thus no connected component of $G \backslash A$) contains more than two elements among a, b, c, d. Moreover, if a connected component C of $G \setminus A$ contains exactly two elements of a, b, c, d, then $N_G(C) \cap A$ is a clique of G separating two elements of a, b, c, d from the two others, which is impossible by Corollary 3 as we supposed $\delta(a, b, c, d) \geq \frac{3}{2}$.

Thus, for every $u \in \{a, b, c, d\} \backslash A$, the connected component C_u of $G \backslash A$ containing u yields a clique $N_G(C_u) \cap A$, which separates u from $\{a, b, c, d\} \setminus \{u\}$. □

As an illustration, one may notice that the central atom in Figure 6 satisfies the property of Lemma 7 with respect to the 4-tuple v_0, v_1, v_2, v_3. Indeed, none of the four vertices is contained into this atom, but each of them is simplicial and can be separated from the three others by its two neighbours.

Moreover, we note that our above result heavily relies on the tree-like structure of the atoms (*e.g* we use a tree-decomposition whose bags are the atoms), and that such a tree-like structure does not exist in general if separators with diameter at least 2 are involved. In fact, our results in this section cannot be extended to decompositions with separators of bounded diameter, as there exist infinitely many graphs that are completely separable with isometric separators of diameter at most 2, and whose hyperbolicity is arbitrarily large [32].

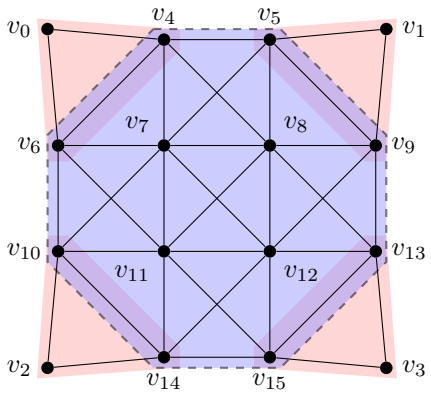


Figure 6: A 2-hyperbolic graph with five atoms: four are 0-hyperbolic, one is 1-hyperbolic.

4.2 Substitution method

From Lemma 7 we can associate a specific atom to a 4-tuple of large hyperbolicity. Four applications of Lemma 5 are then sufficient to prove that the hyperbolicity of this 4-tuple and the hyperbolicity of the atom differ by at most 2. The purpose of this section is to prove that this difference is at most 1. To do this, we refine the results of Section 3.2.

Lemma 8. *Let X be an $(A|B)$-clique-separator of a connected graph G. Given an $(a|b_1, b_2, b_3)$ 4-tuple, write:*

$$\begin{aligned} S_1 &= \mathrm{d}(a, b_1) + \mathrm{d}(b_2, b_3), \\ S_2 &= \mathrm{d}(a, b_2) + \mathrm{d}(b_1, b_3), \\ S_3 &= \mathrm{d}(a, b_3) + \mathrm{d}(b_1, b_2). \end{aligned}$$

Assume w.l.o.g. that $S_1 \geq S_2 \geq S_3$, and let $x_2 \in X$ be such that $\mathrm{d}(a, b_2) = \mathrm{d}(a, x_2) + \mathrm{d}(x_2, b_2) = \mathrm{d}(a, X) + \mathrm{d}(x_2, b_2)$. If $\delta(a, b_1, b_2, b_3) > \delta(x_2, b_1, b_2, b_3)$, then we have

- $S_1 > S_2 = S_3$.
- $\mathrm{d}(a, b_1) = \mathrm{d}(a, x_2) + \mathrm{d}(x_2, b_1)$.

Proof. Recall that by the substitution of Lemma 4, we can assume w.l.o.g. that for every i we have $\mathrm{d}(a, x_i) = \mathrm{d}(a, X) = 1$, where x_i is a vertex of X located on an ab_i-shortest path.

Now, let $\varepsilon_i = \mathrm{d}(x_2, b_i) - \mathrm{d}(x_i, b_i)$. Observe that $\varepsilon_i \in \{0, 1\}$, and that $\varepsilon_i = 0$ if, and only if, x_2 lies on an ab_i-shortest path. In particular we have $\varepsilon_2 = 0$.

Let us denote by S'_i the sum $\mathrm{d}(x_2, b_i) + \mathrm{d}(b_j, b_k)$, where $\{j, k\} = \{1, 2, 3\} \setminus \{i\}$. We aim to exhibit a relation between S_i and S'_i, which would yield in turn a relation between the values $\delta(a, b_1, b_2, b_3)$ and $\delta(x_2, b_1, b_2, b_3)$. First, we have $\mathrm{d}(a, b_i) = \mathrm{d}(x_2, b_i) + 1 - \varepsilon_i$ and so,

$$\begin{aligned} S_i =& \mathrm{d}(a, b_i) + \mathrm{d}(b_j, b_k) \\ =& \mathrm{d}(x_2, b_i) + \mathrm{d}(b_j, b_k) + 1 - \varepsilon_i \\ =& S'_i + 1 - \varepsilon_i \end{aligned}$$

Since we assume here that $S_1 \geq S_2 \geq S_3$ and $\delta(a, b_1, b_2, b_3) > 0$, we have $S'_1 \geq S_1 - 1 \geq \max\{S_2, S_3\} \geq \max\{S'_2, S'_3\}$. More precisely:

- If $S'_2 \geq S'_3$, then $\delta(a, b_1, b_2, b_3) = \delta(x_2, b_1, b_2, b_3) - \varepsilon_1/2 \leq \delta(x_2, b_1, b_2, b_3)$.
- If $S'_3 > S'_2$, then $\varepsilon_3 = 1$ because $S_2 \geq S_3$, which implies $S_2 = S_3$. This, in turn, implies that $\delta(x_2, b_1, b_2, b_3) = (S'_1 - S'_3)/2 = (S_1 - 1 + \varepsilon_1 - S_3)/2 = (S_1 - S_2)/2 - (1 - \varepsilon_1)/2 = \delta(a, b_1, b_2, b_3) - (1 - \varepsilon_1)/2 \leq \delta(a, b_1, b_2, b_3)$.

  In such a case, $\delta(x_2, b_1, b_2, b_3) < \delta(a, b_1, b_2, b_3)$ if, and only if, we have $\varepsilon_1 = 0$, that is x_2 lies on an ab_1-shortest path.

□

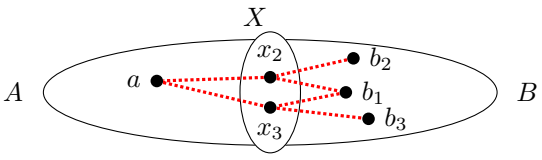


Figure 7: An illustration of the metric property. The dashed lines represent shortest paths.

The metric property of Lemma 8 is illustrated with Figure 7. We use it to extend Lemma 5 as follows.

Lemma 9. *Let a, b, c, d be a 4-tuple of a connected graph G, and two sets $X_a, X_d \subseteq V$ satisfying:*

- *X_a is a $(a|b, c, d)$-clique-separator;*
- *X_d is a $(d|a, b, c)$-clique-separator;*
- *$X_a \backslash X_d$ lies in the same connected component of $G \setminus X_d$ as a, b, c.*

Then there exist $x_a \in X_a$ and $x_d \in X_d$ such that

$$\delta(a, b, c, d) \leq \delta(x_a, b, c, x_d) + 1/2.$$

Proof. Let u_1, u_2, u_3 be such that $\{u_1, u_2, u_3\} = \{b, c, d\}$, $T_1 \geq T_2 \geq T_3$, where $T_i = \mathrm{d}(a, u_i) + \mathrm{d}(u_j, u_k)$ and $\{j, k\} = \{1, 2, 3\} \setminus \{i\}$. We distinguish two cases:

- First, assume that there is $x_a \in X_a$ satisfying $\delta(a, b, c, d) \leq \delta(x_a, b, c, d)$. Then we are done, as Lemma 5 applied to the $(d|x_a, b, c)$-clique separator X_d yields a vertex $x_d \in X_d$ satisfying $\delta(x_a, b, c, d) \leq \delta(x_a, b, c, x_d) + \frac{1}{2}$, hence $\delta(a, b, c, d) \leq \delta(x_a, b, c, x_d) + \frac{1}{2}$.
- Second, assume that no vertex of X_a satisfies the above property. In particular, let $x_a \in X_a$ be such that $\mathrm{d}(a, x_a) = \mathrm{d}(a, X_a)$, and x_a lies on an au_2-shortest path. Note that $\delta(a, b, c, d) \leq \delta(x_a, b, c, d) + \frac{1}{2}$ by Lemma 5 and so, $\delta(a, b, c, d) = \delta(x_a, b, c, d) + \frac{1}{2}$. Moreover, we deduce from Lemma 8 that $T_1 > T_2 = T_3$, and x_a also lies on an au_1-shortest-path. Let $T'_i = \mathrm{d}(x_a, u_i) + \mathrm{d}(u_j, u_k)$. It thus follows that we have:

  $$T'_1 = T_1 - \mathrm{d}(a, X_a), T'_2 = T_2 - \mathrm{d}(a, X_a) \text{ and } T'_3 = T_3 - \mathrm{d}(a, X_a) + 1.$$

  So, we have $T'_1 \geq T'_3 > T'_2$. By the contrapositive of Lemma 8 applied to the 4-tuple $(d|x_a, b, c)$, and noticing that the two least sums amongst T'_1, T'_2, T'_3 are different, there exists a vertex $x_d \in X_d$ satisfying $\delta(x_a, b, c, d) \leq \delta(x_a, b, c, x_d)$.

□

Corollary 10. *Let a, b, c, d be a 4-tuple of a connected graph G satisfying $\delta(a, b, c, d) \geq 3/2$. There exists an atom A of G such that $\delta(a, b, c, d) \leq \delta(G[A]) + 1$*

Proof. The atom A is obtained by applying Lemma 7. For every vertex $u \in \{a, b, c, d\} \setminus A$, let $X_u \subseteq A$ be a clique-separator disconnecting u from $\{a, b, c, d\} \setminus u$. We claim that $A \setminus X_u$ lies in the same connected component of $G \setminus X_u$ as $\{a, b, c, d\} \setminus u$, because otherwise X_u would separate two elements of $\{a, b, c, d\}$ from the two others, which by Corollary 3 implies that $\delta(a, b, c, d) \leq 1$, a contradiction.

One can then apply Lemmas 9 and 5 to find four vertices $a', b', c', d' \in A$ such that $\delta(a, b, c, d) \leq \delta(a', b', c', d') + 1$. Note that the sets $\{a, b, c, d\}$ and $\{a', b', c', d'\}$ are not necessarily disjoint. □

4.3 Additive approximation for hyperbolicity

Theorem 11. *Let $A_1, \ldots, A_l$ be the atoms of a connected graph G. Then:*

$$\max_i \delta(G[A_i]) \leq \delta(G) \leq \max_i \delta(G[A_i]) + 1$$

Proof. As for Theorem 6, the lower-bound of Theorem 11 follows from the fact that the subgraphs $G_i = G[A_i]$ are isometric subgraphs of G. The upper-bound trivially holds when $\delta(G) \leq 1$. We can thus suppose that $\delta(G) \geq 3/2$ and so, that there exist four vertices a, b, c, d such that $\delta(a, b, c, d) = \delta(G) \geq 3/2$. Corollary 10 then yields an atom A such that $\delta(G) \leq \delta(G[A]) + 1$, which proves the second part of our claim. □

Note that the upper-bound is reached by the graph of Figure 6, and by the 1-hyperbolic chordal graph from Figure 1 whose atoms have hyperbolicity 0.

5 Substitution method for an exact computation

As shown with Theorem 11, the maximum hyperbolicity taken over all the atoms may be slightly lesser than the hyperbolicity of the graph. In the two previous sections, we characterized under which situations both values might differ. We now propose to build upon clique-decomposition by replacing the atoms with *substitute graphs*, so that the hyperbolicity of the atoms' substitutes yields the hyperbolicity of the initial graph. This method supposes that the hyperbolicity is at least 1.

Outline of the method. We recall that a *simplicial* vertex is a vertex whose neighborhood induces a complete subgraph. In the spirit of Lemma 4, we add simplicial vertices to the atoms, in order to mimic the maximum $(a|b_1, b_2, b_3)$ 4-tuples that may result from a disconnection. We first introduce our method of substitution in the simpler context of a *single* disconnection. We then focus on technical details related to the implementation.

5.1 Substitute graphs

5.1.1 Basic step: single disconnection

Given a connected graph G, let $V(G) = A \cup B$ such that $X = A \cap B$ is an $(A|B)$-clique separator of G.

- Let $G_A = G[A]$. For every $b \in B \setminus X$, we consider the set of vertices $X_b \subseteq X$ which are at distance $\mathrm{d}_G(b, X)$ from b. For every X_b, we add in G_A a (simplicial) vertex s_{X_b} whose neighborhood is X_b. The resulting graph is named G_A^*.
- Let $G_B = G[B]$. For every $a \in A \setminus X$, we consider the set of vertices $X_a \subseteq X$ which are at distance $\mathrm{d}_G(a, X)$ from a. For every X_a, we add in G_B a (simplicial) vertex s_{X_a} whose neighborhood is X_a. The resulting graph is named G_B^*.

More formally, the *substitute graphs* (or *substitutes* for short) G_A^* and G_B^* of the graphs G_A and G_B with respect to the $(A|B)$-separator X are defined as follows:

Definition 12. *Given a connected graph G, let $V(G) = A \cup B$ such that $X = A \cap B$ is an $(A|B)$-clique separator of G. The* substitute graphs G_A^*, G_B^* *are defined as:*

$$V(G_A^*) = A \cup \{s_{X_b} : \exists b \in B \text{ s.t. } X_b = \arg\min_{x \in X} \mathrm{d}(b, x)\}$$
and $$E(G_A^*) = E(A) \cup \{\{s_{X_b}, x\} : x \in X_b\}$$

$$V(G_B^*) = B \cup \{s_{X_a} : \exists a \in A \text{ s.t. } X_a = \arg\min_{x \in X} \mathrm{d}(a, x)\}$$
and $$E(G_B^*) = E(B) \cup \{\{s_{X_a}, x\} : x \in X_a\}$$

Lemma 13. *Let G be a connected graph satisfying $\delta(G) \geq 1$, and $V(G) = A \cup B$ such that $A \cap B = X$, where X denotes an $(A|B)$-clique separator of G. We define $G_A = G[A]$, $G_B = G[B]$. We have:*

$$\delta(G) = \max\{1, \delta(G_A^*), \delta(G_B^*)\}.$$

Proof. First, we prove that $\delta(G) \leq \max\{1, \delta(G_A^*), \delta(G_B^*)\}$. By construction, the subgraph G_A is an isometric subgraph of G_A^*, and so is G_B for G_B^*. It thus follows that $\max\{\delta(G_A^*), \delta(G_B^*)\} \geq \max\{\delta(G_A), \delta(G_B)\}$. Let us now consider the 4-tuples of G which are separated by the disconnection. We recall that by Corollary 3, the hyperbolicity of any $(a_1, a_2|b_1, b_2)$ 4-tuple of G is bounded by 1. In addition, for every $a \in A \setminus X$, and for every $b_1, b_2, b_3 \in B$, there exists by construction a simplicial vertex a^* of $G_B^* \setminus B$ that is adjacent to $\{x \in X : \mathrm{d}_G(a, x) = \mathrm{d}_G(a, X)\}$; hence, $\delta(a, b_1, b_2, b_3) = \delta(a^*, b_1, b_2, b_3) \leq \delta(G_B^*)$ by Lemma 4. The proof is symmetric for $b \in B \setminus X$ and $a_1, a_2, a_3 \in A$ and it yields the expected result.

To prove $\delta(G) \geq \max\{1, \delta(G_A^*), \delta(G_B^*)\}$, let us consider w.l.o.g. an arbitrary 4-tuple of G_B^*. If such a 4-tuple does not contain a simplicial vertex of $G_B^* \setminus B$, then we are done as it exists in the isometric subgraph G_B of G. If it contains exactly one simplicial vertex $a^* \in G_B^* \setminus B$, then by construction we can replace a^* with a vertex $a \in A \setminus X$, satisfying $N(a^*) = \{x \in X : \mathrm{d}_G(a, x) = \mathrm{d}_G(a, X)\}$; this operation does not modify the hyperbolicity of the 4-tuple by Lemma 4. Finally, if it contains at least two simplicial vertices of $G_B^* \setminus B$, then it is an $(a_1, a_2|b_1, b_2)$ 4-tuple of G_B^* and so, the hyperbolicity is bounded by 1. □

We emphasize that some simple rules can be applied to reduce the size of the substitute graphs. In particular, we can remove the pendant vertices which may be added in the construction.

5.1.2 Extension to the atoms

The substitution operation can be naturally extended to the whole clique-decomposition, by mimicking each step of it and applying the basic substitution operation that we describe above at each of these steps. We formalize it by first introducing the following definition of an atom tree.

Definition 14 ([6, 7, 34]). *Let G be a connected graph. An* atom tree *of G is a labeled binary rooted tree T, satisfying the following recursive definition:*

- *if G is prime w.r.t. clique-decomposition, then T is reduced to a node labeled with V;*
- *otherwise, the root of T is labeled with a clique-minimal separator X, and there exists a connected component C of $G \setminus X$ satisfying:*
  - *$N_G(C) \setminus C = X$;*
  - *the left child of the root is labeled with $A = C \cup X$, which does not contain any clique-minimal separator;*
  - *and the right subtree of the root is an atom tree of $G \setminus C$.*

In order to prevent any confusion, the reader has to notice that an atom tree is *not* a tree-decomposition (as defined in Section 4.1). In fact, an atom tree can be seen as the trace of some execution of the algorithm of [34, 45] to compute the clique-decomposition. Indeed, it is proved in [34] that in an atom tree, the leaves are in bijective correspondance with the atoms of the graph. Given a *fixed* atom tree, this yields a natural total ordering of the atoms by increasing depth. We now follow this ordering to construct the substitutes of the atoms from the atom tree. There are as many steps for our substitution method as there are atoms in the graph.

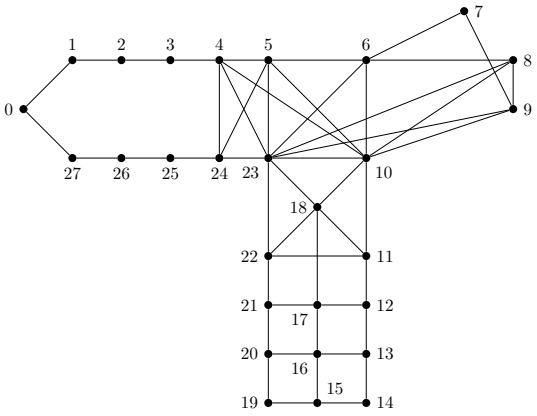
- Starting from $H_1 = G$, we disconnect the first atom A_1 by using the clique-minimal separator X_1 from the atom tree. Applying the substitution operation of Definition 12 to $A = A_1$ and $B = V(G) \setminus (A_1 \setminus X_1)$, we obtain two substitute graphs: $G^*_{A_1}$ which substitutes A_1, and another one denoted by $H_2 = G^*_B$.
- After $i - 1$ steps, $i \in \{2, \ldots, l-1\}$, we constructed the substitute graphs of atoms $A_1, \ldots, A_{i-1}$, plus an additional graph H_i. The graph H_i contains $G[\bigcup_{j \geq i} A_j]$, to which were added simplicial vertices during the previous steps. By using the clique-minimal separator X_i from the atom tree, we disconnect the graph H_i, and we apply the substitution operation of Definition 12, this time to the set A equal to $C_A \cup X$ where C_A is the connected component of $H_i \setminus X$ which intersects A_i, and to $B = V(H_i) \setminus (A \setminus X_i)$. We replace H_i with the two substitute graphs, one containing the atom A_i and being its substitute, the other being denoted by $H_{i+1} = G^*_B$.
- We finally stop at the l^{th} step, and we set H_l is the substitute graph of the *last* atom A_l.

Figure 8 illustrates this process. The numbers reported in Table 8i illustrate the interest of our pre-processing method for the computation of the hyperbolicity. Indeed, the graph G of Figure 8a has 28 nodes and so, 20 475 4-tuples, while the sum of the numbers of 4-tuples in the graphs G^*_i (Figures 8c–8h) is 1 800. We thus significantly reduce the search space. Moreover, a simple cutting rule allows us to reduce the number of 4-tuples to consider to 1 575. To do so, we first order the graphs G^*_i by decreasing diameters, then we iteratively compute the hyperbolicity of these graphs in this order, and we stop exploration as soon as the diameter of a graph G^*_j is smaller than twice the largest value of δ computed so far.

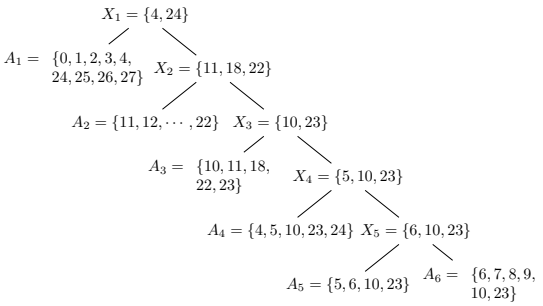
5.2 Implementation and complexity analysis

5.2.1 Precomputation step and updates

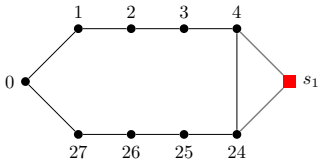
We first focus on some computational tasks that have to be repeatedly executed at each step of our substitution method. In this section, we provide a high-level description for an implementation of these tasks.



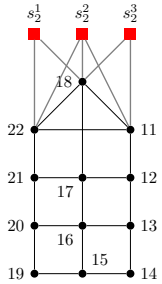
(a) G



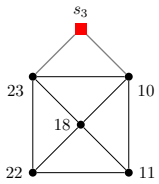
(b) Atom tree of G



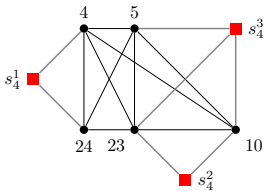
(c) G_1^* built from A_1



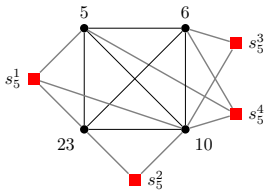
(d) G_2^* built from A_2



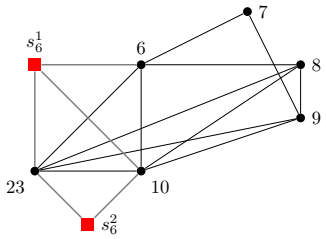
(e) G_3^* built from A_3



(f) G_4^* built from A_4



(g) G_5^* built from A_5



(h) G_6^* built from A_6

	n	m	D	δ	4-tuples
G	28	50	9	2	20 475
G_1^*	10	11	5	2	210
G_2^*	15	25	5	2	1 365
G_3^*	6	10	2	1	15
G_4^*	8	16	3	1/2	70
G_5^*	8	16	2	1/2	70
G_6^*	8	16	3	1	70

(i) Main characteristics

Figure 8: A connected graph G (Figure 8a), an atom tree of the graph (Figure 8b), the substitute of the atoms of G (Figures 8c–8h), and the characteristics of these graphs (Table 8i).

Computation of the distances We heavily rely in our implementation on constant time access to the distance between some vertex (possibly added by our construction) and the vertices of some clique-minimal separator. It is thus essential that we update the distance matrix of the graph quickly when we replace atoms by their substitutes.

Lemma 15. *Let G be a connected graph. We can embed in quadratic time the distance matrix of G into a data structure, supporting:*

- $O(1)$ *access to the distance between a vertex contained in a clique-minimal separator and any other;*
- $O(1)$ *updates when a simplicial vertex is added by our substitution method.*

Proof. The gist of such a structure is Lemma 4. Let X be an $(A|B)$-clique separator of G, and s be a simplicial vertex added in the substitute graph of $G[B]$. Let $a \in A \setminus X$ satisfy $N(s) = \{x \in X : \mathrm{d}_G(a,x) = \mathrm{d}_G(a,X)\}$. Then we have for every $b \in B$, $\mathrm{d}(b,s) = \mathrm{d}_G(a,b) - \mathrm{d}_G(a,X) + 1$.

It thus follows that once the substitution of a with s has been completed, we only need to remember the association of s with a and an offset, so that we can compute the distances in the substitute graphs. The offset can be computed in constant time by picking a neighbor of the simplicial vertex s, as it is the distance $\mathrm{d}(a,X)$ between this neighbor and the vertex a which s substitutes. Finally, since there are $l = O(n)$ steps for our substitution method, and that no more than $O(n)$ new simplicial vertices are added at each step, a quadratic-size array is sufficient to store all the pairs $(a, \mathrm{d}(a,X_a))$. □

Note that the data structure of Lemma 15 does not support the computation of distances between two vertices added by our construction. We can safely ignore this drawback, as we do not need to compute such distances in our method.

Determining the vertex sets of the substitute graphs Another problem is to compute, at step i of the algorithm, the connected component containing the next atom A_i to deal with. Determining the connected components of a graph is linear-time computable. However, as we detailed in Section 5.1.2, here we have to extract the component from a graph $H_i \neq G$, possibly containing more edges than G due to the addition of simplicial vertices at previous steps. Thus it may result in an $\Omega(m)$-time complexity by using the classical algorithm for this problem. Instead, we propose a method to construct the component incrementally, starting from A_i and adding simplicial vertices at every step $1 \leq j \leq i-1$.

Lemma 16. *Let G be a connected graph, T be an atom tree of G, and $A_1, \ldots, A_l$ be its atoms ordered according to their depth in T. We denote by $H_1 = G, H_2, \ldots, H_l$ the sequence of l graphs that are computed by our process, each H_i being decomposed into H_{i+1} and the substitute graph of the i^{th} atom by applying the substitution method of Definition 12. For every (simplicial) vertex $s_i \in H_{i+1} \setminus H_i$, we can compute the index j such that s_i belongs to the substitute graph of the j^{th} atom, in total $O(n|X_i|)$-time.*

Proof. Let $s_i \in H_{i+1} \setminus H_i$ be a simplicial vertex. By construction (Section 5.1.2), we have $N(s_i) \subseteq X_i \subseteq V(G)$. Therefore, if s_i belongs to the substitute graph of the j^{th} atom, $j > i$, then it holds that $N(s_i)$ intersects the connected component containing $A_j \setminus X_j$ |in the graph $H_j \setminus X_j$|. In such case since every vertex in the component either belongs to the atom or is simplicial and not in $V(G)$, then it follows that s_i has a neighbour in $A_j \setminus X_j$. Conversely, if $s_i \in V(H_j)$ and $N(s_i) \cap (A_j \setminus X_j) \neq \emptyset$, then s_i is in the same connected component as $A_j \setminus X_j$ in the graph $H_j \setminus X_j$, hence s_i belongs to the substitute graph of the j^{th} atom. So, at every step of our substitution method, if a simplicial vertex is added by our construction, we consider the

minimum index j such that $A_j \setminus X_j$ contains a neighbour of the vertex, and we update the vertex set of the substitute graph of the j^{th} atom by adding this new vertex into it. Since only $O(n)$ vertices are added at step i, and that their neighborhood is contained into X_i, the $O(n|X_i|)$-time complexity follows. □

We now focus on some algorithmic aspects of the substitution operation (Section 5.1.1).

5.2.2 Applying simplification rules

In this section, we will assume that the distance matrix of the graph is given. Given the set $B \setminus X$, it is straightforward to compute in $O(n|X|)$-time all of the subsets $X_b = \{x \in X : \mathrm{d}_G(b,x) = \mathrm{d}_G(b,X)\}$, for every $b \in B \setminus X$. A naive implementation would then consist in adding a simplicial vertex for every b and by doing so, we would lose all the benefit of the separation in terms of size of the graphs. We now define rules in order to avoid this worst-case in some situations, in order to decrease the number of simplicial vertices to add in the substitute graphs. The goal of this section is to give hints on an efficient way to implement these rules.

Partition refinement techniques Indeed, it may happen that $X_b = X_{b'}$ for some pair $b, b' \in B$, and in such a case we wish to add only one simplicial vertex in the substitute graph G^*_A.

To do that efficiently, we will use the well-known *partition refinement* techniques (*e.g.* see [27, 39]). Given a partition $\mathcal{P}$ of a set V, and a subset $S \subseteq V$ called the *pivot*, the partition refinement of $\mathcal{P}$ w.r.t. S consists in replacing every group V_i of $\mathcal{P}$ by the non-empty groups amongst $V_i \cap S$ and $V_i \cap \bar{S}$. This can be achieved in $O(|S|)$-time, up to the precomputation of an appropriate data structure in linear $O(|V|)$-time.

We deduce from this standard technique the following result:

Lemma 17. *Let G be a connected graph given by its distance matrix, and $X \subseteq V(G)$. We define the relation $\equiv_X$ over the set $V(G) \setminus X$ as*

$$u \equiv_X v \text{ iff } \{x \in X : \mathrm{d}_G(u,x) = \mathrm{d}_G(u,X)\} = \{x \in X : \mathrm{d}_G(v,x) = \mathrm{d}_G(v,X)\}.$$

The equivalence classes of $\equiv_X$ can be computed in $O(n|X|)$-time.

Proof. We start from the partition $\mathcal{P} = \{V \setminus X\}$ which we refine successively for every $x \in X$ with the set $\{u : u \in V \setminus X \text{ s.t. } \mathrm{d}_G(u,x) = \mathrm{d}_G(u,X)\}$. The total cost is $O(\sum_{x \in X} |N_{G_X}(x)|) = O(n|X|)$. □

5.2.3 Complexity analysis

Finally, to determine the time complexity of our substitution method, we will use the following result:

Lemma 18 ([6])**.** *Let G be a connected graph, and $A_1, \ldots, A_l$ be its atoms. Then $\sum_i |A_i| \leq n+m$.*

Corollary 19. *The substitute of the atoms of a connected graph G can be computed in $O(nm)$-time.*

Proof. Let T be an atom tree of G, and $A_1, \ldots, A_l$ be the atoms of the graph. For every i, let X_i be the clique-minimal separator of G labeling the father node of leaf A_i in T. By Definition 14, $X_i \subseteq A_i$.

We first precompute the distance matrix of G in $O(nm)$-time, then we embed it in quadratic-time into the data structure of Lemma 15. Also, for every i, we initialize the vertex set of the

i^{th} substitute graph with the atom A_i. We then apply each step of our substitution method sequentially.

We can easily check that at step i, there are at most $O(n)$ vertices to consider. By Lemma 17, this allows us to compute the simplicial vertices to add at this step in $O(n|X_i|)$-time. By Lemma 16, we can then update the vertex sets of the substitute graphs with the same time complexity. Since at most $O(n)$ vertices are added at this step, we finally update the distances in $O(n)$-time by Lemma 15.

Consequently, our modified clique-decomposition can be computed in $O(n\sum_i |X_i|)$-time, that is in $O(n\sum_i |A_i|) = O(nm)$-time by Lemma 18. □

6 Hyperbolicity of outerplanar graphs

Last, we deal with an algorithmic application of the substitution method of Section 5. We use it here to provide a *linear-time* algorithm for computing the hyperbolicity of a large class of graphs.

A *planar graph* is a graph drawable in the Euclidean plane so that edges may only intersect at their endpoints. It is *outerplanar* if it stays planar whenever one adds a universal vertex to it. Equivalently, a graph is outerplanar if it is drawable in the Euclidean plane so that edges may only intersect at their endpoints, and all the vertices lie on a common face which is called the *outerface*. Such a drawing is furthermore called an outerplanar embedding, and it can be computed in linear-time [43]. Outerplanar graphs are a minor-closed graph class, and a graph is outerplanar if, and only if, it is K_4-minor-free and $K_{2,3}$-minor-free (see [42] for other characterizations).

All cycles are outerplanar graphs, and in particular all atoms of an outerplanar graph are cycles. The hyperbolicity of cycles is determined by the following formula:

Lemma 20 ([46, 14])**.** *Cycles of order $4p+\varepsilon \geq 3$, with $p \geq 0$ and $\varepsilon \in \{0,1,2,3\}$, are $(p-1/2)$-hyperbolic when $\varepsilon = 1$, and p-hyperbolic otherwise.*

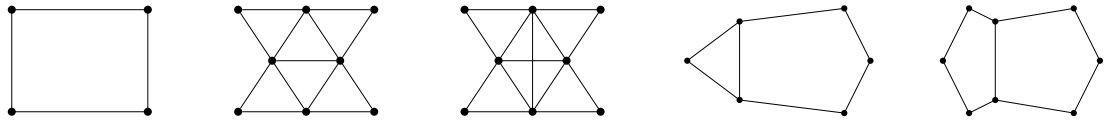
We will use Lemma 20 to prove the main result of this section (*e.g.* Theorem 27). Our proofs heavily rely on the notion of *weak dual* [2].

Definition 21. *Let G be a biconnected outerplanar graph. The weak dual of G is a tree T_G equal to the intersection graph of the atoms of G. Two adjacent nodes of T_G correspond to atoms which share a single edge.*

Note that a weak dual is nothing else than a tree-decomposition whose bags are the atoms of an outerplanar graph. If G is biconnected, then starting from an outerplanar embedding of the graph, we construct it by removing from the dual of the graph the universal vertex corresponding to the outerface (see Figure 10 for an example).

6.1 Outerplanar graphs with small hyperbolicity

As observed in Section 5, our substitution method for an exact computation of hyperbolicity requires the hyperbolicity of the graph to be at least 1. To overcome this drawback, we first characterize in this section outerplanar graphs that are $\frac{1}{2}$-hyperbolic. Note that we only consider biconnected graphs, as the hyperbolicity of a graph is the maximum hyperbolicity taken over all its biconnected components, and the biconnected components of a graph are computable in linear-time [44].



(a) C_4 (b) H_1 (c) H_2 (d) H_3 (e) H_4

Figure 9: Characterization of 5-chordal $\frac{1}{2}$-hyperbolic graphs, in terms of forbidden isometric subgraphs.

Proposition 22. *A biconnected outerplanar graph is $\frac{1}{2}$-hyperbolic if, and only if, either it is isomorphic to C_5, or it is chordal and it does not contain the graph of Figure 9b as a subgraph. Furthermore, these conditions can be checked in linear-time.*

Proof. Let G be a $\frac{1}{2}$-hyperbolic outerplanar biconnected graph. By Lemma 20, the graph C_5 is $\frac{1}{2}$-hyperbolic, and we now assume G is not isomorphic to C_5. As every induced cycle of G is isometric, the induced cycles of G are exactly its atoms. As a result, we have by Lemma 20 that G only has induced cycles of length 3 or 5. Moreover, Wu and Zhang prove in [46] that a 5-chordal graph is $\frac{1}{2}$-hyperbolic if, and only if, it does not contain any graph of Figure 9 as an isometric subgraph[1].

By the hypothesis, G is C_4-free and so, it does not contain the graph of Figure 9a as an isometric subgraph. Moreover, we claim that G is C_5-free, as otherwise it would contain the graph of Figure 9d, or the graph of Figure 9e, as an isometric subgraph. Thus G has to be chordal. Also, we check that G cannot contain the graph of Figure 9c, as it is not outerplanar and being outerplanar is a hereditary property. Consequently, G is $\frac{1}{2}$-hyperbolic if, and only if, it is chordal and it does not contain the graph of Figure 9b as an isometric subgraph.

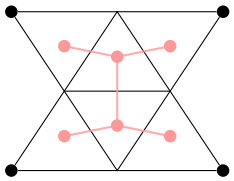


Figure 10: The forbidden subgraph of Figure 9b, and its characterization in the weak dual.

Finally, recall that a graph is outerplanar if, and only if, it is K_4-minor-free and $K_{2,3}$-minor-free. Let H_1 be the graph of Figure 9b. We claim that every induced subgraph of G that is isomorphic to H_1 is isometric, as otherwise it would yield a $K_{2,3}$-minor for G. In the same way, every subgraph of G that is isomorphic to H_1 is an induced subgraph of G, as otherwise it would yield a K_4-minor for G. So, G is $\frac{1}{2}$-hyperbolic if, and only if, it is chordal and it does not contain H_1 as a subgraph. Being chordal can be checked in linear-time [40], and a chordal outerplanar graph contains H_1 as a subgraph if, and only if, there are two adjacent vertices of degree 3 in its weak dual (see Figure 10). □

6.2 Substitute graphs of cycles

As we constrain ourselves to outerplanar graphs, recall that the atoms are exactly the induced cycles of the graph. Clearly, a clique-separator contained into a cycle is either a cut-point (in

[1]The characterization of [46] is composed of six forbidden isometric subgraphs, but the sixth one is actually 6-chordal.

which case our substitution method never adds a simplicial vertex), or it is an edge-separator (in which case our substitution method might only add a single vertex, which has to be adjacent to both endpoints of the edge). Substitute graphs of cycles thus fall into the following subclass of outerplanar graphs:

Definition 23. *A sunshine graph is a biconnected outerplanar graph, containing a dominating induced cycle C satisfying every vertex not in C is a simplicial vertex of degree 2.*

Cycles are a subclass of such graphs, and two other examples of sunshine graphs are given in Figure 11. Moreover, we remark that an edge in a dominating cycle may be the neighborhood of at most one simplicial vertex out of the cycle, as otherwise it would yield a $K_{2,3}$-minor for the graph. It can be noted in addition that except for the particular case of a diamond, there exists a unique dominating cycle in a sunshine graph, and every induced cycle which is not dominating is a triangle. Thus, if G is a sunshine graph and C is a dominating cycle of G, then we have by Theorem 11 that $\delta(C) \leq \delta(G) \leq \delta(C) + 1$. This difference can actually be decreased by $\frac{1}{2}$ as follows.

Lemma 24. *Let G be a sunshine graph, and C be a dominating cycle of G. Then we have:*

$$\delta(C) \leq \delta(G) \leq \delta(C) + \frac{1}{2}$$

Proof. By Theorem 11, we have $\delta(C) \leq \delta(G) \leq \delta(C) + 1$. So, our aim is to prove that no 4-tuple of G has a hyperbolicity greater than $\delta(C) + \frac{1}{2}$. By contradiction, let a, b, c, d be such that $\delta(a, b, c, d) = \delta(C) + 1$.

We arbitrarily orient the cycle C. For every $u \in \{a, b, c, d\} \setminus C$, we denote by $e_u = \{x_u, y_u\}$ the edge of C induced by its neighbours, where x_u denotes the head of the edge w.r.t. the orientation. Observe that for every $u, v \in \{a, b, c, d\} \setminus C$, we have $\mathrm{d}(u, v) = 2 + \min\{\mathrm{d}(x_u, y_v), \mathrm{d}(x_v, y_u)\} = 1 + \mathrm{d}(x_u, x_v) = 1 + \mathrm{d}(y_u, y_v)$.

We then claim that there is exactly one vertex amongst a, b, c, d which belongs to the cycle C. Indeed, not all of a, b, c, d belong to C as the 4-tuple has a hyperbolicity greater than $\delta(C)$. Furthermore, at least three of them are not in C, as otherwise we would have $\delta(a, b, c, d) \leq \delta(C) + \frac{1}{2}$ by Lemmas 5 and 9. On the other hand, if none of them is in C, then we can check that $\delta(a, b, c, d) = \delta(x_a, x_b, x_c, x_d) \leq \delta(C)$, a contradiction. So, the claim is proved.

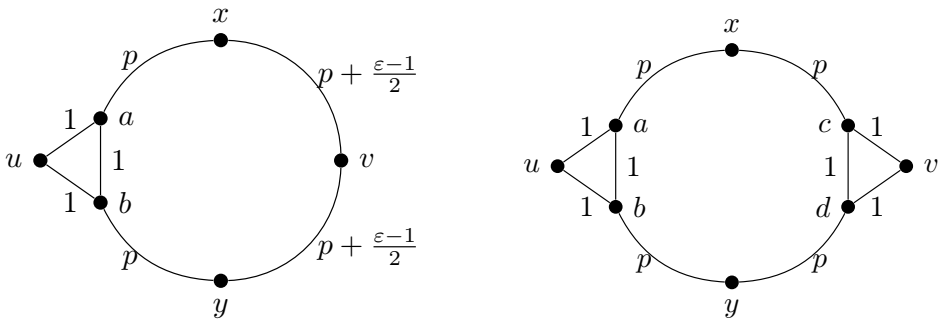
Finally, assume w.l.o.g. that $a \in C$. In such a case, we can check that $\delta(a, b, c, d) \leq \delta(a, x_b, x_c, x_d) + \frac{1}{2} \leq \delta(C) + \frac{1}{2}$. A contradiction. □

Once the upper-bound for hyperbolicity is refined as above, the hyperbolicity of a sunshine graph can be computed in *linear-time*, using the following characterization.

Lemma 25. *Let G be a sunshine graph, and C be a dominating cycle for G of length $4p+\varepsilon \geq 3$, with $p \geq 0$ and $\varepsilon \in \{0, 1, 2, 3\}$. Assuming $G \setminus C$ is nonempty we have:*

- *if ε is odd, then $\delta(G) = \delta(C) + \frac{1}{2}$;*
- *if $\varepsilon = 2$, then $\delta(G) = \delta(C) + \frac{1}{2}$ if, and only if, there exist two simplicial vertices out of C which are diametrically opposed;*
- *finally, if $\varepsilon = 0$, then $\delta(G) = \delta(C)$.*

Proof. Recall that by the previous Lemma 25, we have $\delta(G) \leq \delta(C) + \frac{1}{2}$. Thus we only focus on finding 4-tuples u, v, x, y of hyperbolicity (at least) this value, and we choose one, if any, maximizing $|C \cap \{u, x, v, y\}|$. In the sequel, write $S_1 = \mathrm{d}(u, v) + \mathrm{d}(x, y)$, $S_2 = \mathrm{d}(u, x) + \mathrm{d}(v, y)$ and $S_3 = \mathrm{d}(u, y) + \mathrm{d}(v, x)$. We will assume in addition that $S_1 \geq S_2 \geq S_3$.



(a) n odd (b) $n = 4p + 2$

Figure 11: Substitute graphs of the atoms of an outerplanar graph.

Case ε odd Equivalently, we have $\varepsilon \in \{1, 3\}$. In such a case, we have $\delta(C) = p + \frac{\min\{0,\varepsilon-2\}}{2}$ by Lemma 20. Figure 11a exhibits a 4-tuple u, v, x, y satisfying:

$$S_1 = (2p + \frac{\varepsilon+1}{2}) + (2p + \min\{1, \varepsilon - 1\}) = 4p + \frac{\varepsilon+1}{2} + \min\{1, \varepsilon - 1\}$$
$$S_2 = (p + 1) + (p + \frac{\varepsilon-1}{2}) = 2p + \frac{\varepsilon+1}{2}$$
$$S_3 = S_2$$

Hence, this 4-tuple has hyperbolicity $p + \frac{\min\{1,\varepsilon-1\}}{2} = \delta(C) + \frac{1}{2}$.

Case $\varepsilon = 2$ In such a case, we have $\delta(C) = p$ by Lemma 20. We assume w.l.o.g. that $u \notin C$, and we claim that it implies $v \notin C$. Indeed, by the metric property of Lemma 8, and noticing that $S_1 \geq S_2 \geq S_3$, the vertex v has to be at equal distance l of both neighbours of u, as otherwise u could be replaced with one of its two neighbours, contradicting the maximality of $|C \cap \{u, x, v, y\}|$. Hence $v \notin C$ is impossible, as it would yield the length of C is $2l + 1 = 4p + 2$. It thus follows that $v \notin C$, and the length of C is in fact $2(l - 1) + 2 = 2l$, yielding $l = 2p + 1$.

Conversely, assume that there exist two simplicial vertices u, v that are diametrically opposed in G. We choose the 4-tuple u, x, v, y as in Figure 11b, and it satisfies:

$$S_1 = (2p + 2) + (2p + 1) = 4p + 3$$
$$S_2 = 2(p + 1) = 2p + 2$$
$$S_3 = S_2$$

So, we have $\delta(u, v, x, y) = p + \frac{1}{2} = \delta(C) + \frac{1}{2}$.

Case $\varepsilon = 0$ Another application of Lemma 20 yields $\delta(C) = p$. Assuming $u \notin C$, we deduce as for the previous case that $v \notin C$, and v is at equal distance $l = 2p$ from both neighbours of u. Thus, C is partitioned by the neighborhoods of u and v, in the same way as in Figure 11b, into two paths of length $l - 1 = 2p - 1$. Furthermore, since the diameter of C is $2p$, those paths are geodesics of the cycle. By the proof of Lemma 24, at least one vertex $z \in \{x, y\}$ amongst the 4-tuple has to be in C so that $\delta(u, v, x, y) > \delta(C)$. As a result, by picking the geodesic containing z we get $\mathrm{d}(u, z) + \mathrm{d}(v, z) = 2 + (l - 1) = 2p + 1$. So, we have $\min\{\mathrm{d}(u, z), \mathrm{d}(v, z)\} \leq p$, contradicting

the fact that $\delta(u,v,x,y) > \delta(C)$, because we have by [41] that $\delta(u,v,x,y) \leq \min\{\mathrm{d}(u,z), \mathrm{d}(v,z)\}$. To sum up, we always have $\delta(G) = \delta(C)$ in such a case. □

6.3 Applying the substitution method in linear-time

Recall that our substitution method consists in constructing a substitute graph for every atom of the graph, then computing the hyperbolicity of every substitute graph, and finally taking the maximum over these hyperbolicity values and 1. In the case of an outerplanar graph, we provided in the previous section a characterization yielding a linear-time computation of the hyperbolicity of the substitute graphs of the atoms (*e.g.* Lemma 25). What now remains to prove is that the substitute graphs can also be computed in linear-time.

Lemma 26. *Let G be an outerplanar biconnected graph. The substitute graphs of the atoms of G can be computed in linear-time.*

Proof. First, we construct in linear-time an outerplanar embedding for G, then we construct from it the weak dual T_G of G. Let $C_1, \ldots, C_l$ be the atoms of G. We root T_G on an atom C_1, which is an induced cycle. Note that $(T_G; C_1)$ naturally yields an atom tree, as defined in Definition 14: starting from a rooted tree $(T; C)$ initialized with $(T_G; C_1)$, we iterate as long as T is not empty, and at each step we disconnect all the atoms being the leaves of T sequentially. Such an ordering is enough to prove the correctness of the following algorithm.

- For every i, we initialize C_i^* with C_i.
- We start a depth-first search from the root, and for every visited atom C_i, once its subtree has been visited, we consider all its sons C_j, naming $e_{i,j}$ the edge-separator $C_i \cap C_j$. We add in C_i^* a simplicial vertex whose neighborhood is $e_{i,j}$ if, and only if, there is a vertex in C_j^* which is at equal distance to both endpoints of $e_{i,j}$. This is equivalent to have either the length of C_j is odd, or there is a simplicial vertex in $C_j^* \setminus C_j$ whose neighborhood is the edge diametrically opposed to $e_{i,j}$ in C_j.
- Finally, we start a breadth-first search from the root and for every visited atom $C_i \neq C_1$, we consider its parent atom, denoted by C_k, naming $e_{i,k}$ the edge-separator that it shares with it. As before, we add in C_i^* a simplicial vertex whose neighborhood is $e_{i,k}$ if, and only if, either the length of C_k is odd, or there is a simplicial vertex in $C_k^* \setminus C_k$ whose neighborhood is the edge diametrically opposed to $e_{i,k}$ in the atom C_k.

Note that for an atom, the depth-first search here is used to compute the simplicial vertices resulting from the disconnection of its sons, whereas the breadth-first search is used to compute the single vertex resulting from its own disconnection, if any. Using an atom tree as defined above, it can be checked that the resulting $C_1^*, \ldots, C_l^*$ are the substitute graphs of the atoms of G. □

Figure 12 shows the substitute graphs resulting from the application of the substitution method to a biconnected outerplanar graph.

We finally conclude with the following theorem.

Theorem 27. *The hyperbolicity of a connected outerplanar graph is computable in linear-time.*

Proof. We can safely assume G to be biconnected by [44]. By [4, 28], G is 0-hyperbolic if, and only if, G is a clique. If it is not, then by Proposition 22, we can check whether it is $\frac{1}{2}$-hyperbolic in linear-time.

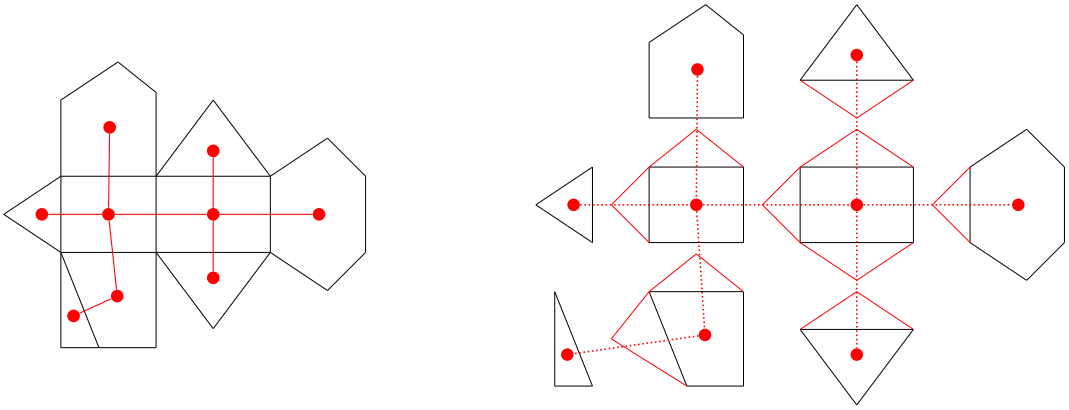


Figure 12: An application of the substitution method to an outerplanar graph.

From now on, assume $\delta(G) \geq 1$. By Lemma 26, we can compute the substitute graphs of the atoms of G in linear-time. We can thus conclude by Lemma 13 (*i.e.* the correctness of our substitution method), as these substitute graphs are sunshine graphs and their hyperbolicity is linear-time computable by Lemma 25. □

7 Experimental evaluation

We report in this section on experiments performed with our substitution methodology on the graphs of five collaboration networks. This way, we aim to evaluate the computation time of the substitutes on some empirical graphs, and to better understand the factors impacting their size.

7.1 Datasets

We apply the algorithm presented in Section 5 to the collaboration networks of five different scientific communities [35], namely:

- `ca-AstroPh`, for the astrophysics community;
- `ca-CondMat`, for the condensed matter physics community;
- `ca-GrQc`, for the general relativity and quantum cosmology community;
- `ca-HepPh`, for the high energy physics-phenomenology community;
- and `ca-HepTh`, for the high energy physics-theory community.

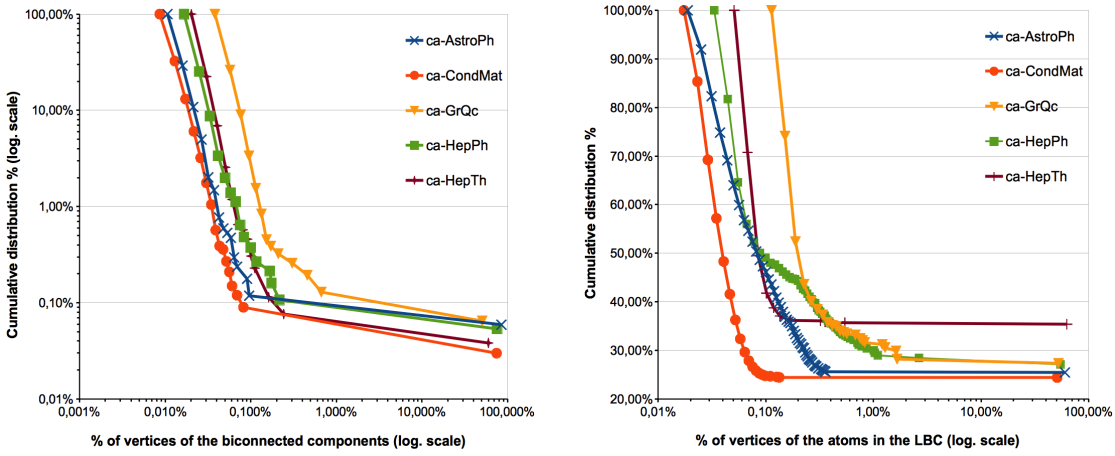
In the `ca-*` graphs, nodes represent scientists and edges represent collaborations (i.e., co-authoring a paper). These graphs are interesting to analyze the behavior of our algorithm, and the size of their substitute graphs, because they have many cliques of various sizes. Indeed, a paper co-authored by k scientists induces a clique of size k in the graph. Furthermore, the number of co-authors per papers varies from one community to another. Therefore, we will observe different results in terms of the size of their substitute graphs, despite these graphs share many properties (see [35]).

7.2 Empirical results

We modified the clique-decomposition algorithm of [7], to implement the substitution method that we presented in Section 5. We used it here to compute, for every graph, the substitute of each atom of the decomposition.

Decomposition into biconnected components We observed that all of the five graphs are composed of one largest biconnected component, that we call LBC. The component includes from 50% to 84.85% of all the vertices. This can be observed from the cumulative distribution of the size of the biconnected components in Figure 13a. The cumulative number of components is given as a percentage of the total number of biconnected components, and the size of the components as a percentage of the total number of vertices in the graph. We noticed that all the biconnected components but the LBC are small: only covering at most 1% of the vertices.

Clearly, the smallest biconnected components can be safely ignored for the computation of hyperbolicity, provided that a) their number of vertices is less than 4, or that b) their diameter is smaller than two times the hyperbolicity of the LBC, which is always the case for these graphs (see [14]). Thus, we now focus on the clique-decomposition of the LBC, and on its resulting substitute graphs.



(a) Biconnected components

(b) Atoms in the LBC

Figure 13: Cumulative distributions of the size of the biconnected components (Figure 13a) and of the atoms in the LBC of each graph (Figure 13b).

Clique-decomposition We plotted in Figure 13b the cumulative distribution of the size of the atoms of the LBC. The cumulative number of atoms is given as a percentage of the total number of atoms and the size of the atoms as a percentage of the total number of vertices in the LBC. Again, for all of the graphs, we observed one largest atom, that we call the LA. This atom includes from 50% to 60% of all the vertices, and all the other atoms only represent a small fraction of the overall vertices. In the worst case (`ca-HepPh`), all the atoms but the LA solely cover 2.65% of the vertices of the graph.

Moreover, like for the smallest biconnected components and as reported in [14], the substitute graphs of the smallest atoms can be safely ignored for the computation of hyperbolicity. As a

result, the only component of the graphs to deal with for computing their hyperbolicity is the substitute graph of the LA. We will denote it by LS in the sequel.

Size of the substitute graphs As explained in Section 5, the size of the LS depends on both the initial size of the LA and the number of added simplicial vertices. We have reported in Table 1 the original size n of each graph, the size n_B of its LBC, the size n_{LA} of the LA, and the size n_{LS} of the largest substitute. We have then computed the percentage R_{LA} of vertices that have been removed from the LBC to obtain the LA, that is $R_{LA} = \frac{n_B - n_A}{n_B}$. We observe a significant reduction rate R_{LA}, varying from 37.40% to 49.22%. We have also computed the reduction rate R_{LS} of the LS with respect to the LBC, that is $R_{LS} = \frac{n_B - n_{LS}}{n_B}$. We observe that this reduction rate falls between 11.22% and 20.84%. It indicates that the substitution method adds many simplicial vertices to the LA, when constructing the LS, despite the simplification rules that are presented in Section 5.2.2.

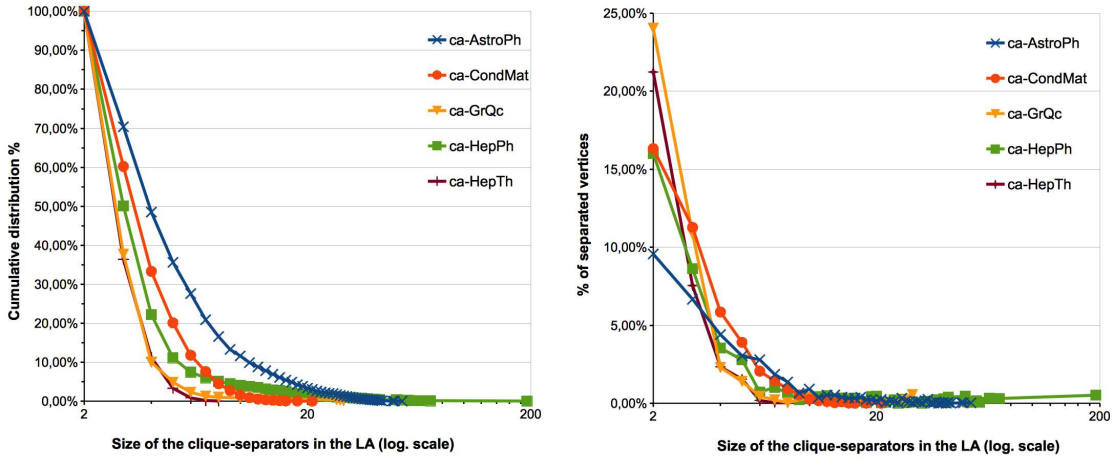
Instance name	n	n_B	n_{LA}	n_{LS}	R_{LA}	R_{LS}	Cost	Time
`ca-CondMat`	23 133	17 234	8 751	13 643	49.22%	20.84%	28.39%	672
`ca-GrQc`	5 242	2 651	1 386	2 107	47.72%	20.52%	27.20%	5
`ca-HepPh`	12 008	9 025	4 925	7 170	45.43%	20.55%	24.88%	167
`ca-AstroPh`	18 772	15 929	9 561	13 407	39.98%	15.83%	24.14%	679
`ca-HepTh`	9 877	5 898	3 692	5 236	37.40%	11.22%	26.18%	53

Table 1: Characteristics of the collaboration networks. The size of the graph is given as n, the size of the LBC as n_b, the size of the LA as n_{LA} and the size of the LS as n_{LS}. The percentage of vertices removed from the LBC to obtain the LA is given as R_{LA}, the reduction rate is $R_{LS} = \frac{n_B - n_{LS}}{n_B}$, and the percentage of vertices in the LBC, representing the addition of simplicial vertices, is given as *Cost*. Finally, the computation time of the substitution method, denoted by $Time$, is given in seconds.

We reported in Table 1 as *Cost* the percentage of vertices in the LBC representing the addition of new simplicial vertices. We first observe the impact of the addition of new simplicial vertices on `ca-CondMat` which has the largest reduction rates R_{LS} and R_{LA}. Despite a R_{LA} of 49.22%, `ca-CondMat` has almost the same reduction rate R_{LS} as `ca-HepPh` and `ca-GrQc` |ranging from 20.55% to 20.84%|. This is the consequence of the addition of new simplicial vertices which represents 28.39% of the size of its LBC, whereas for the `ca-HepPh` graph it goes up to only 24.88%. A similar behavior is observed between `ca-AstoPh` and `ca-CondMat`: even though their R_{LA} differ on 9.25%, the difference of their reduction rate R_{LS} finally falls to 5%. This results from the addition of 4.24% less simplicial vertices in `ca-AstoPh` than in `ca-CondMat`. As an extremal case when looking at the `ca-HepPh` graph, the difference of 2.29% with the R_{LA} of `ca-GrQc` is even *compensated* by the addition of 2.32% less simplicial vertices, resulting in a R_{LS} smaller than for `ca-GrQc`. However, as another extremal case when comparing `ca-HepTh` to `ca-CondMat`, we notice that a large difference of 11.82% between their respective R_{LA} is only marginally decreased by the addition of 2.21% less simplicial vertices for `ca-HepTh` than for `ca-CondMat`. The substitution method results in 9.61% more vertices for `ca-HepTh` than for `ca-CondMat`. We thus conclude that the impact of n_{LA} and of the number of new simplicial vertices on the final size n_{LS} differs greatly depending on the graph.

7.3 Decomposition analysis

Having noticed the heterogeneous results of our empirical section, we are now analyzing in more details the properties causing the asymmetry between the various `ca-*` graphs.



(a) Clique-separators in the LA

(b) Separated vertices from the LA

Figure 14: Cumulative distribution of the size of the clique-separators in the LA (Figure 14a) and percentage of separated vertices as a function of the size of the clique-separators in the LA (Figure 14b).

Clique-decomposition We first analyzed the composition of the LA in terms of clique-separators. Let us denote by $\mathcal{X}_{LA} = \{X_1, \ldots, X_l\}$ the clique-minimal separators that are contained into the LA, disconnecting the atoms $\mathcal{A}_{LA} = \{A_1, \ldots, A_l\}$ from the LA, sequentially. We emphasize that there might be other atoms in the graph than the LA and those in $\mathcal{A}_{LA}$. But such atoms, if any, do not overlap the LA. Said differently, the set $\mathcal{A}_{LA}$ represents the neighbourhood of the LA in the *atom graph*, as it is defined in [6].

We plotted in Figure 14a the cumulative distribution of the size of the clique-separators in the LA as a percentage of the total number of clique-separators. By doing so, we observed smaller clique-separators for the `ca-HepTh` and `ca-CondMat` graphs, with a maximum size of 8 and 21, respectively, than for the three others graphs `ca-GrQc`, `ca-AstroPh` and `ca-HepPh`, having clique-separators of maximum size 42, 53 and 192, respectively. Also, we reported in Table 2 that the ratio $R_{|\mathcal{X}_{LA}|}$ of the number of clique-minimal separators $|\mathcal{X}_{LA}|$ over the size n_{LA} of the LA, varies from 0.39 for `ca-AstroPh` to 0.54 for `ca-CondMat`. To sum up, there are more clique-separators in `ca-CondMat` than in `ca-AstropPh`, but there are larger clique-separators in `ca-AstropPh` than in `ca-CondMat`.

To complete our measurements, we related the size of clique-separators with the proportion of vertices that are disconnected by them from the LA. We reported in Table 2 as $\alpha_1 = n_B - n_{LA}$ the total number of vertices separated from the LA in the LBC, as $\alpha_2 = |V(\mathcal{A}_{LA} \setminus \mathcal{X}_{LA})|$ the number of vertices in the subset of atoms $\mathcal{A}_{LA}$ which are not in the LA, and as $|\mathcal{X}_{LA}|$ the number of clique-minimal separators in the LA. Finally, we computed the fraction $\Delta_1 = \frac{\alpha_1 - \alpha_2}{n_B}$, quantifying the percentage of vertices that are neither contained into the LA, nor in any of the

atoms in $\mathcal{A}_{LA}$. We reported as $\Delta_2 = R_{LA} - \Delta_1$ the fraction of vertices in some atom of $\mathcal{A}_{LA}$, hence those that are *directly* separated from the LA.

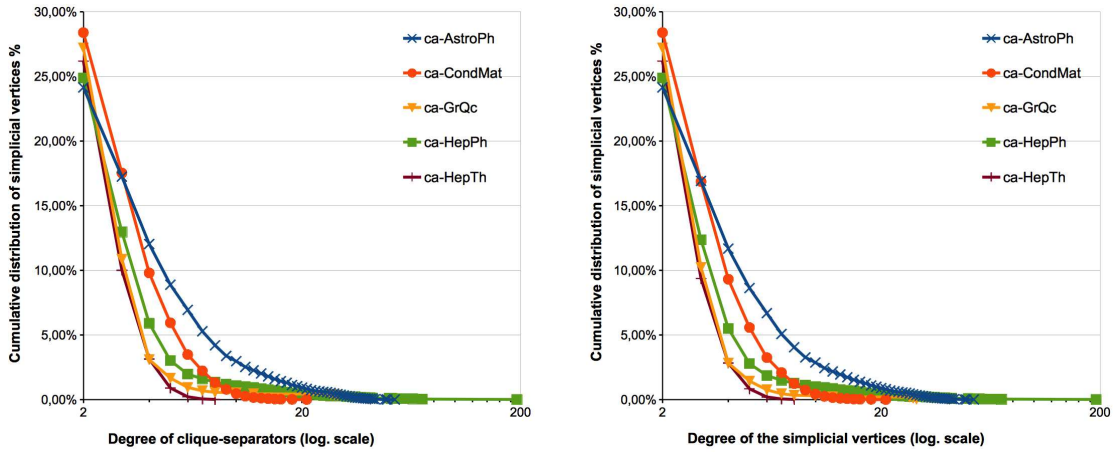
Our results put in evidence that most of the vertices are either contained into the LA, or into some other atom intersecting the LA. Other vertices comprise around 2.88% and 7.03% of the overall vertices. Moreover, as shown with Figure 14b, where we plotted the percentage of separated vertices as a function of the size of clique-minimal separators, smaller clique-separators of size ≤ 5 are responsible for a significant part (w.r.t. Δ_2) of the vertices disconnected from the LA in `ca-CondMat` (37.34% of vertices over 49.22%), whereas in `ca-AstroPh` they solely disconnect 23.67% over 39.98% of vertices. This difference is not balanced with clique-separators of larger size, even though these ones disconnect 13.43% of vertices in `ca-AstroPh`, while only 5.67% in `ca-CondMat`. A comparison of `ca-CondMat` with `ca-HepPh` yielded similar results. In contrast, for the graphs `ca-GrQc` and `ca-HepTh`, we notice that 6.71% and 4.91% more vertices, respectively, than in `CondMat`, are disconnected by edge-separators. But the rest of the clique-minimal separators, of larger size, only disconnect 16.67% and 11.70% of the vertices, respectively, whereas 26.70% of them are separated in `ca-CondMat`. Therefore, most of the difference for the final size of the substitute graph LS comes from the number of vertices that are disconnected by clique-minimal separators of *small* size.

Instance name	α_1	α_2	$\lvert\mathcal{X}_{LA}\rvert$	$R_{\mathcal{X}_{LA}}$	Δ_1%	Δ_2%
`ca-CondMat`	8 483	7 413	4 702	0.54	6.21%	43.01%
`ca-GrQc`	1 265	1 079	698	0.5	7.03%	40.69%
`ca-HepPh`	4 100	3 727	2 166	0.44	4.13%	41.3%
`ca-AstroPh`	6 368	5 910	3 715	0.39	2.88%	37.1%
`ca-HepTh`	2 206	1 942	1 506	0.41	4.47%	32.93%

Table 2: Distribution of clique-minimal separators, and of the vertices disconnected from the LA.
The total number of vertices separated from the LA in the LBC is given as $\alpha_1 = n_B - n_{LA}$, and the number of disconnected vertices being present in the subset of neighbouring atoms $\mathcal{A}_{LA}$ as $\alpha_2 = |V(\mathcal{A}_{LA} \setminus \mathcal{X}_{LA})|$.
Also, the number of clique-minimal separators in the LA is given as $|\mathcal{X}_{LA}|$.
We quantify the percentage of vertices that are neither contained into the LA nor in any of the atoms in $\mathcal{A}_{LA}$ as $\Delta_1 = \frac{\alpha_1 - \alpha_2}{n_B}$; the fraction of vertices in some atom of $\mathcal{A}_{LA}$ that are *directly* separated from the LA is equal to $\Delta_2 = R_{LA} - \Delta_1$.

Substitute construction In order to validate our previous assumption, namely that the largest number of simplicial vertices are connected to the smallest clique-separators, we plotted in Figure 15a the cumulative number of simplicial vertices connected to the LA, normalized by the size of the LBC, as a function of the size of the clique-minimal separators. In particular, note that for each graph, such a summation is equal to the value given as `Cost` in Table 1. By looking only at clique-minimal separators of size two and three, the proportions of simplicial vertices for the graphs `ca-CondMat`, `ca-GrQc`, `ca-HepPh`, `ca-AstroPh` and `ca-HepTh` respectively, represent 65.49%, 88.63%, 76.26%, 50.16% and 88.02% respectively, of the total number of simplicial vertices connected to the LA. Thus it highlights the importance of clique-minimal separators of small size, to which a large proportion of simplicial vertices are connected to.

Let us also remark by comparing Figure 15a to Figure 15b that almost all simplicial vertices



(a) Simplicial vertices connected to the LA as a function of the size of the clique-separators

(b) Degree distribution of the simplicial vertices connected to the LA

Figure 15: Cumulative number of simplicial vertices connected to the LA normalized by the size of the LBC as a function of the size of the clique-separators to which they are connected (Figure 15a), cumulative degree distribution of the simplicial vertices connected to the LA normalized by the size of the LBC (Figure 15b)

are connected to clique-separators *of the same size*. In the worst case (`ca-GrQc`), there are no more than 0.75% of the simplicial vertices whose degree differs from the others. Most of these simplicial vertices have degree two. Hence, the final proportion of simplicial vertices, given in Table 1 as `Cost`, mostly depends on the size distribution of the clique-separators in the graphs. Also, there is a worst-case variation of only 4.25% -between `ca-CondMat` and `ca-AstroPh`- in the proportion of simplicial vertices in our graphs, which allows us to make relative comparisons between them. Especially we are interested in comparing the proportion of simplicial vertices of small degree (less than four). Such a proportion represents, for `ca-CondMat`, `ca-GrQc`, `ca-HepPh`, `ca-AstroPh` and `ca-HepTh` respectively, a percentage of 18.59%, 24.1%, 18.97%, 12.11% and 23.04% respectively, of the simplicial vertices in total. To sum up:

- when comparing `ca-AstroPh` to `ca-CondMat`: even if the former has 2.23% more simplicial vertices with degree more than three, this is compensated by the 6.48% less simplicial vertices of degree less than four, which results in overall to 4.25% less simplicial vertices in `ca-AstroPh` than in `ca-CondMat`. The same happens when comparing `ca-AstroPh` to the remaining graphs. The lower number of simplicial vertices with degree less than four *always* compensates its larger number of simplicial vertices that are connected to clique-separators of larger size.

- when comparing `ca-CondMat` to `ca-GrQc` and `ca-HepTh`: the two latter graphs respectively have 5.51% and 4.45% more simplicial vertices with degree smaller than four. However, they respectively have 6.7% and 6.66% less simplicial vertices that are adjacent to clique-separators of size greater than three. As a result, there are 1.19% less simplicial vertices in `ca-GrQc`, and 2.21% less simplicial vertices in `ca-HepTh`, respectively, than in `ca-CondMat`.

- finally, when comparing `ca-CondMat` to `ca-HepPh`: we observe quite similar numbers of

simplicial vertices that are connected to clique-separators of degree smaller than four. They respectively represent 18.59% and 18.97% of the simplicial vertices in total. Again, the main difference comes from the proportion of simplicial vertices with degree higher than three, with 5.91% more simplicial vertices in `ca-CondMat` than in `ca-HepPh`, resulting in 3.51% less vertices in `ca-HepPh`.

Our experiments on the `ca-*` graphs have highlighted the importance of the size of the largest atom, from which the largest substitute graph is always constructed. On the one hand in our examples, the three graphs `ca-CondMat`, `ca-GrQc` and `ca-HepPh`, which have the largest ratio R_{LA}, also have the smallest ratio R_{LS}. But on the other hand, the cost of adding new simplicial vertices to construct the LS can greatly compensate a smaller ratio R_{LA}, as it is the case for `ca-HepPh`. Indeed this graph has a smaller ratio R_{LA} than `ca-GrQc`, but it finally ends up with a larger ratio R_{LS}. Finally, we observed on every graph of our experimentation that the degree distribution of the simplicial vertices follows the size distribution of the clique-separators. In particular, simplicial vertices of degree two and three always represent from 50.16% to 88.63% of the simplicial vertices in total.

8 Conclusion

In this work, we proved a tight relation between the hyperbolicity of a graph and the maximum hyperbolicity taken over all its atoms. Our result subsumes the one in [11], as it directly implies that chordal graphs are 1-hyperbolic. Also, it implies that two graph classes extending chordal graphs, namely 2-chordal graphs [36, 37] and clique-separable graphs [25], have a bounded hyperbolicity.

In addition of our main result, we deduced from its proof a general substitution method, allowing us to modify the atoms at no extra-cost than the clique-decomposition; for graphs with hyperbolicity at least one, the maximum hyperbolicity taken over all the resulting graphs is exactly the hyperbolicity of the graphs, but the graphs to be considered may have a larger size than the atoms. Experiments suggest that the final size of the substitute graphs is mostly related with the number of clique-minimal separators of small size, and the disconnections resulting from them. We successfully applied this substitution method to outerplanar graphs, providing in this case a linear-time algorithm for computing the hyperbolicity. We let open whether the same can be done for other classes of graphs.

Finally, in the spirit of [2], it would be of interest to take advantage of the linear-time algorithm for outerplanar graphs, in order to yield an efficient computation of the hyperbolicity of planar graphs. Part of our future work will also consist in finding other graph decompositions which are applicable to the computation of this parameter.

References

[1] M. Abu-Ata and F. F. Dragan. Metric tree-like structures in real-life networks: an empirical study. Technical Report arXiv:1402.3364, ArXiv, 2014.

[2] B. Baker. Approximation algorithms for NP-complete problems on planar graphs. *Journal of the ACM (JACM)*, 41(1):153–180, 1994.

[3] H.-J. Bandelt and V. Chepoi. Metric graph theory and geometry: a survey. *Contemporary Mathematics*, 453:49–86, 2008.

[4] H.-J. Bandelt and H. Mulder. Distance-hereditary graphs. *Journal of Combinatorial Theory, Series B*, 41(2):182–208, 1986.

[5] S. Bermudo, J. Rodríguez, J. Sigarreta, and J.-M. Vilaire. Gromov hyperbolic graphs. *Discrete Mathematics*, 313(15):1575–1585, 2013.

[6] A. Berry, R. Pogorelcnik, and G. Simonet. Efficient clique decomposition of a graph into its atom graph. Technical Report RR-10-07, LIMOS, Aubière, France, Mar. 2010.

[7] A. Berry, R. Pogorelcnik, and G. Simonet. An introduction to clique minimal separator decomposition. *Algorithms*, 3(2):197–215, 2010.

[8] H. L. Bodlaender. Discovering treewidth. In *31st Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, volume 3381 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2005.

[9] M. Boguñá, F. Papadopoulos, and D. V. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1(62):1–18, Oct. 2010.

[10] J. A. Bondy and U. Murty. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[11] G. Brinkmann, J. H. Koolen, and V. Moulton. On the hyperbolicity of chordal graphs. *Annals of Combinatorics*, 5(1):61–69, 2001.

[12] J. Chakerian and S. Holmes. Computational tools for evaluating phylogenetic and hierarchical clustering trees. *Journal of Computational and Graphical Statistics*, 21(3):581–599, 2012.

[13] V. Chepoi, F. F. Dragan, B. Estellon, M. Habib, and Y. Vaxès. Notes on diameters, centers, and approximating trees of δ-hyperbolic geodesic spaces and graphs. *Electronic Notes in Discrete Mathematics*, 31:231–234, 2008.

[14] N. Cohen, D. Coudert, and A. Lancin. Exact and approximate algorithms for computing the hyperbolicity of large-scale graphs. Rapport de recherche RR-8074, Inria, Sept. 2012.

[15] D. Coudert and G. Ducoffe. On the recognition of C_4-free and 1/2-hyperbolic graphs. Research Report RR-8458, INRIA, Jan. 2014.

[16] W. H. Cunningham. Decomposition of directed graphs. *SIAM Journal on Algebraic Discrete Methods*, 3(2):214–228, 1982.

[17] P. de La Harpe and E. Ghys. *Sur les groupes hyperboliques d'après Mikhael Gromov*, volume 83. Progress in Mathematics, 1990.

[18] M. Didi Biha, B. Kaba, M.-J. Meurs, and E. SanJuan. Graph decomposition approaches for terminology graphs. In *MICAI 2007: Advances in Artificial Intelligence*, volume 4827 of *Lecture Notes in Computer Science*, pages 883–893. Springer, 2007.

[19] R. Diestel. *Graph theory, graduate texts in mathematics, vol. 173*. Springer, Heidelberg, 1997.

[20] Y. Dourisboure and C. Gavoille. Tree-decompositions with bags of small diameter. *Discrete Mathematics*, 307(16):2008–2029, 2007.

[21] F. Dragan. Tree-like structures in graphs: A metric point of view. In *39th International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, volume 8165 of *Lecture Notes in Computer Science*, pages 1–4. Springer, 2013.

[22] A. Dress, K. Huber, J. Koolen, V. Moulton, and A. Spillner. *Basic Phylogenetic Combinatorics*. Cambridge University Press, Cambridge, UK, Dec. 2011.

[23] H. Fournier, A. Ismail, and A. Vigneron. Computing the gromov hyperbolicity of a discrete metric space. Technical Report arXiv:1210.3323, ArXiv, Oct. 2012.

[24] T. Gallai. Transitiv orientierbare graphen. *Acta Mathematica Hungarica*, 18(1):25–66, 1967.

[25] F. Gavril. Algorithms on clique separable graphs. *Discrete Mathematics*, 19(2):159–165, 1977.

[26] M. Gromov. Hyperbolic groups. In S. Gersten, editor, *Essays in Group Theory*, volume 8 of *Mathematical Sciences Research Institute Publications*, pages 75–263. Springer, New York, 1987.

[27] M. Habib, C. Paul, and L. Viennot. Partition refinement techniques: An interesting algorithmic tool kit. *International Journal of Foundations of Computer Science*, 10(02):147–170, 1999.

[28] E. Howorka. On metric properties of certain clique graphs. *Journal of Combinatorial Theory, Series B*, 27(1):67–74, 1979.

[29] E. A. Jonckheere and P. Lohsoonthorn. Geometry of network security. In *American Control Conference*, volume 2, pages 976–981, Boston, MA, USA, 2004. IEEE.

[30] B. Kaba, N. Pinet, G. Lelandais, A. Sigayret, and A. Berry. Clustering gene expression data using graph separators. *In silico biology*, 7(4):433–452, 2007.

[31] W. S. Kennedy, O. Narayan, and I. Saniee. On the hyperbolicity of large-scale networks. Technical Report arXiv:1307.0031, ArXiv, 2013.

[32] J. H. Koolen and V. Moulton. Hyperbolic bridged graphs. *European Journal of Combinatorics*, 23(6):683–699, 2002.

[33] F. Le Gall. Faster algorithms for rectangular matrix multiplication. In *IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 514–523, New Brunswick, NJ, USA, 2012. IEEE.

[34] H.-G. Leimer. Optimal decomposition by clique separators. *Discrete Mathematics*, 113(1):99–123, 1993.

[35] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1):1–41, Mar. 2007.

[36] S. McCullough. 2-chordal graphs. In *Contributions to Operator Theory and its Applications*, pages 143–192. Springer, 1988.

[37] S. McCullough. Minimal separators of 2-chordal graphs. *Linear algebra and its applications*, 184:187–199, 1993.

[38] K. Olesen and A. Madsen. Maximal prime subgraph decomposition of bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(1):21–31, 2002.

[39] R. Paige and R. Tarjan. Three partition refinement algorithms. *SIAM Journal on Computing*, 16(6):973–989, 1987.

[40] D. Rose, R. Tarjan, and G. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on computing*, 5(2):266–283, 1976.

[41] M. A. Soto Gómez. *Quelques propriétés topologiques des graphes et applications à internet et aux réseaux*. PhD thesis, Univ. Paris Diderot (Paris 7), 2011.

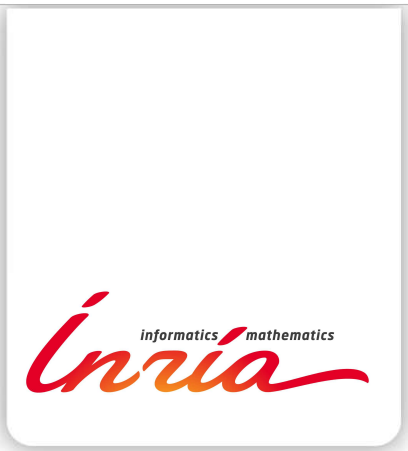
[42] M. Sysło. Characterizations of outerplanar graphs. *Discrete Mathematics*, 26(1):47–53, 1979.

[43] R. Tamassia and I. Tollis. Planar grid embedding in linear time. *IEEE Transactions on Circuits and Systems*, 36(9):1230–1234, 1989.

[44] R. E. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.

[45] R. E. Tarjan. Decomposition by clique separators. *Discrete Mathematics*, 55(2):221–232, 1985.

[46] Y. Wu and C. Zhang. Hyperbolicity and chordality of a graph. *The Electronic Journal of Combinatorics*, 18(1):P43, 2011.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399